11-30-01 PCT

FORM PTO-1390 (REV 10-95) U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

ATTORNEY'S DOCKET NUMBER: 0380-P02746US0

U.S. APPLICATION NO. (If known, see 37 CFR 1.5): 09/980217

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/GB00/02072	30 May 2000	28 May 1999

TITLE OF INVENTION: POLYKETIDES AND THEIR SYNTHESIS

APPLICANT(S) FOR DO/EO/US: LEADLAY, Peter Francis et al.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. [ ] This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. [X] A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. [ ] is transmitted herewith (required only if not transmitted by the International Bureau).
   b. [X] has been transmitted by the International Bureau.
   c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [ ] A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. [X] Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. [ ] are transmitted herewith (required only if not transmitted by the International Bureau).
   b. [ ] have been transmitted by the International Bureau.
   c. [ ] have not been made; however, the time limit for making such amendemnts has NOT expired.
   d. [X] have not been made and will not be made.
8. [ ] A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. [ ] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11. to 16. below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [ ] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment.
    [ ] A SECOND or SUBSEQUENT preliminary amendment.
14. [ ] A substitute specification.
15. [ ] A change of power of attorney and/or address letter.
16. [X] Other items or information:

    Copy of Form PCT/IB/308

09/980217 . 032606

JC10 Rec'd PCT/PTO 28 NOV 2001

U.S. APPLICATION NO (if known, see 37 CFR 1.5) 09/980217	INTERNATIONAL APPLICATION NO PCT/GB00/02072	ATTORNEY'S DOCKET NUMBER 0380-P02746US0

17. [X] The following fees are submitted:	CALCULATIONS PTO USE ONLY	
BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):		
Search Report has been prepared by the EPO or JPO		
International preliminary examination fee paid to USPTO (37 CFR 1.482)		
No international preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2))		
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO		
International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4)		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 890.00	
Surcharge of $130.00 for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$ 130.00	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total claims	47 - 20 =	17	X 18.00	$ 306.00	
Independent claims	16 - 3 =	13	X 84.00	$ 1,092.00	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ 280.00	$ 0.00	
	TOTAL OF ABOVE CALCULATIONS =			$ 2,418.00	
Reduction of 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also by filed (Note 37 CFR 1.9, 1.27, 1.28).				$ 1,209.00	
	SUBTOTAL =			$ 1,209.00	
Processing fee of $130.00 for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(f)). +				$ 0.00	
	TOTAL NATIONAL FEE =			$ 1,209.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). $40.00 per property +				$ 0.00	
	TOTAL FEES ENCLOSED =			$ 1,209.00	
				Amount to be: refunded	$
				charged	$

a. [X] A check in the amount of $ 1,209.00 to cover the above fees is enclosed.

b. [ ] Please charge my Deposit Account No. _____ in the amount of $_____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [X] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 04-1406. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

HAGAN, Patrick J.
Dann Dorfman Herrell and Skillman, P.C.
1601 Market Street, Suite 720
Philadelphia, Pennsylvania 19103
United States of America

Patrick J. Hagan
SIGNATURE

Patrick J. Hagan
NAME

27,643
REGISTRATION NUMBER

THE UNITED STATES PATENT AND TRADEMARK OFFICE

United States Serial No.	:	Not yet assigned
International Application No.	:	PCT/GB00/02072
International Filing Date	:	30 May 2000
Inventor(s)	:	Peter Francis Leadlay et al.
Title	:	POLYKETIDES AND THEIR SYNTHESIS

Suite 720
1601 Market Street
Philadelphia, PA 19103-2307
(215) 563-4100 (telephone)
(215) 563-4044 (facsimile)
Our File: 0380-P02746US0

Box PCT
U.S. Patent and Trademark Office
P.O. Box 2327
Arlington, VA 22202

PRELIMINARY AMENDMENT

Dear Sir:

Before calculation of the filing fee, please amend the claims of the above-referenced patent application, as follows:

In the Claims:

Please amend claims 1, 8-15, 17-23, 26, 29, 33, 35, 37, 38 and 42 as follows:

1. (Amended) A DNA sequence which is selected from the group consisting of (a) at least part of the sequence set out in the appended sequence listing; and (b) a variant of a sequence (a) which encodes a polypeptide which is at least 80%, identical with the corresponding peptide as set out in table II; provided that it is not a sequence encoding all or part of the polypeptide consisting of amino acids 1-920 encoded by *mon AI* as set out in table II.

8. (Amended) A DNA sequence according to claim 1 encoding any one or more of the domains as set out in Table I or a variant or part thereof.

9. (Amended) A DNA sequence according to claim 1 which has a length of at least 30 bases.

10. (Amended) A recombinant cloning or expression vector comprising a DNA sequence according to claim 1.

11. (Amended) A transformant host cell which has been transformed to contain a DNA sequence according to claim 1 and which is capable of expressing a corresponding polypeptide.

12. (Amended) A hybridisation probe which is a DNA sequence according to claim 1.

13. (Amended) A method of detecting a PKS cluster comprising using a probe according to claim 12 to detect a PKS cluster, optionally followed by isolation of the detected cluster.

14. (Amended) A method of detecting genes comprising using a probe according to claim 12 which encodes at least part of a polypeptide having a known function to detect genes encoding polypeptides having analogous function.

15. (Amended) A method according to claim 14 wherein the polypeptide of known function is AT of module 5 or the regulatory protein encoded by *mon RI*.

17. (Amended) A method of detecting the presence of a gene cluster which governs the synthesis of a polyether, which comprises using a probe according to claim 16, and optionally isolating a gene cluster detected thereby.

18. (Amended) A method of detecting a gene comprising using a probe according to claim 12 which comprise a polynucleotide which binds specifically to a gene responsible for levels of activity of the monensin gene cluster, for detecting an

analogous gene in a gene cluster for biosynthesis of another polyketide, optionally followed by a step of manipulating the gene detected thereby to alter the level of expression of said other polyketide.

19. (Amended) A method according to claim 18 wherein the gene is a regulatory gene, resistance gene or thioesterase gene.

20. (Amended) A method of expressing a heterologous gene in *S. cinnamonesis* comprising inserting said gene so that it is expressed under the control of the *mon RI* gene or variant and a monensin promoter.

21. (Amended) A method of expressing a polyketide other than monensin which includes using a portion of the monensin gene cluster encoding a polypeptide having chain terminating activity, comprising at least one of *mon AIX* and *mon AX* or a mutant, allele or other variant thereof encoding a polypeptide having chain terminating activity, to effect chain release of said polyketide other than monensin.

22. (Amended) A method of synthesising a polyketide other than monensin which includes using a portion of the monensin gene cluster encoding a polypeptide having carbon-carbon double bond isomerase activity comprising at least one of

mon BI and *mon BII* or a mutant, allele or other variant thereof having isomerase activity to provide a desired stereochemical outcome in the synthesis of said polyketide other than monensin.

23. (Amended) A polypeptide encoded by a portion of the monensin gene cluster, comprising at least one portion selected from *mon BI* and *mon BII* or a mutant, allele or other variant thereof, having carbon-carbon double bond isomerase activity, or at least one of *mon AIX* and *mon AX* or a mutant, allele or other variant thereof having chain terminating activity.

26. (Amended) A method for the biosynthesis of a polyketide other than monensin which comprises using a portion of the monensin gene cluster encoding a peptide having epoxidase or cyclase activity, to provide a said activity in the biosynthesis of said polyketide other than monensin.

29. (Amended) A process according to claim 27 wherein the starter unit also includes an AT_q domain derived from an AT domain which is naturally associated with the KS domain.

33. (Amended) A DNA sequence according to claim 30 wherein said loading module is adapted to load a starter unit other than a starter unit normally received by the adjacent extension module.

35. (Amended) A polyketide synthase encoded by the DNA sequence of claim 30.

37. (Amended) A vector containing a DNA sequence of claim 30.

38. (Amended) A transformant cell transformed to contain a DNA sequence of claim 30.

42. (Amended) A method of producing monensin comprising culturing the organism of claim 41.

Please add new claims 46 and 47 as follows:

46. (New) A DNA sequence according to claim 1 which is a variant of a sequence (a) which encodes a peptide which is at least 90% identical with the corresponding peptide as set out in table II.

47. (New) A DNA sequence according to claim 1 which has a length of at least 60 bases.

REMARKS

The purpose of this Preliminary Amendment is to eliminate multiple claims dependencies, revise claims which, due to their form, do not comply with current U.S. Patent and Trademark Office practice, and to present additional claims directed to preferred embodiments of the invention.

The foregoing amendments do not introduce new matter into the present application, and, therefore, should be entered without objection.

Early and favorable consideration of the present application is respectfully requested.

Respectfully submitted,

Patrick J. Hagan

Patrick J. Hagan
Reg. No. 27,643
Attorney for Applicant

PJH:ksk

MARKED-UP COPY OF THE CLAIMS

1. (Amended) A DNA sequence which is selected from the group consisting of (a) at least part of the sequence set out in the appended sequence listing; [or] and (b) a variant of a sequence (a) which encodes a polypeptide which is at least 80%, [preferably at least 90%], identical with the corresponding peptide as set out in table II; provided that it is not a sequence encoding all or part of the polypeptide consisting of amino acids 1-920 encoded by *mon AI* as set out in table II.

8. (Amended) A DNA sequence according to [any preceding] claim 1 encoding any one or more of the domains as set out in Table I or a variant or part thereof.

9. (Amended) A DNA sequence according to [any preceding] claim 1 which has a length of at least 30[, preferably at least 60,] bases.

10. (Amended) A recombinant cloning or expression vector comprising a DNA sequence according to [any preceding] claim 1.

11. (Amended) A transformant host cell which has been transformed to contain a DNA sequence according to [any of claims 1-9] claim 1 and which is capable of expressing a corresponding polypeptide.

12. (Amended) A hybridisation probe which is a DNA sequence according to [any of claims 1-9] claim 1.

13. (Amended) A method of detecting a PKS cluster comprising using [Use of] a probe according to claim 12 to detect a PKS cluster, optionally followed by isolation of the detected cluster.

14. (Amended) A method of detecting genes comprising using [Use of] a probe according to claim 12 which encodes at least part of a polypeptide having a known function to detect genes encoding polypeptides having analogous function.

15. (Amended) A method [Use] according to claim 14 wherein the polypeptide of known function is AT of module 5 or the regulatory protein encoded by *mon RI*.

17. (Amended) [Use of a probe according to claim 16 in a] A method of detecting the presence of a gene cluster which governs the synthesis of a polyether, which comprises using a probe according to claim 16, and optionally isolating a gene cluster detected thereby.

18. (Amended) [Use of] A method of detecting a gene comprising using a probe according to claim 12 which comprise a polynucleotide which binds specifically to a gene

responsible for levels of activity of the monensin gene cluster, [in a method of] for detecting an analogous gene in a gene cluster for biosynthesis of another polyketide, optionally followed by a step of manipulating the gene detected thereby to alter the level of expression of said other polyketide.

19. (Amended) A method [Use] according to claim 18 wherein the gene is a regulatory gene, resistance gene or thioesterase gene.

20. (Amended) A method of expressing a heterologous gene in *S. cinnamonesis* comprising inserting said gene so that it is expressed under the control [Use] of the *mon RI* gene or variant and a monensin promoter [to control expression of a heterologous gene in *S. cinnamonensis*].

21. (Amended) A method of expressing a polyketide other than monensin which includes using [Use of] a portion of the monensin gene cluster encoding a polypeptide having chain terminating activity, [preferably] comprising at least one of *mon AIX* and *mon AX* or a mutant, allele or other variant thereof encoding a polypeptide having chain terminating activity, to effect chain release of [a peptide] said polyketide other than monensin.

22. (Amended) A method of synthesising a polyketide other than monensin which includes using [Use of] a portion of the monensin gene cluster encoding a polypeptide having carbon-carbon double bond isomerase activity[, preferably] comprising at least one of *mon BI* and *mon BII* or a mutant, allele or other variant thereof having isomerase activity to provide a desired stereochemical outcome in the synthesis of [a] said polyketide other than monensin.

23. (Amended) A polypeptide encoded by a portion of the monensin gene cluster, [preferably] comprising at least one [of] portion selected from *mon BI* and *mon BII* or a mutant, allele or other variant thereof, having carbon-carbon double bond isomerase activity, or at least one of *mon AIX* and *mon AX* or a mutant, allele or other variant thereof having chain terminating activity.

26. (Amended) A method for the biosynthesis of a polyketide other than monensin which comprises using [Use of] a portion of the monensin gene cluster encoding a peptide having epoxidase or cyclase activity, [preferably comprising *mon CI* or *mon CII* or a mutant, allele or other variant thereof encoding a polypeptide having epoxidase or cyclase activity] to provide a said activity in the biosynthesis of [a polypeptide] said polyketide other than monensin.

29. (Amended) A process according to claim 27 [or claim 28] wherein the starter unit also includes an AT_q domain derived from an AT domain which is naturally associated with the KS domain.

33. (Amended) A DNA sequence according to claim 30[, 31 or 32] wherein said loading module is adapted to load a starter unit other than a starter unit normally received by the adjacent extension module.

35. (Amended) A polyketide synthase encoded by the DNA sequence of [any of claims 30-34] claim 30.

37. (Amended) A vector containing a DNA sequence of [any of claims 30-34] claim 30.

38. (Amended) A transformant cell transformed to contain a DNA sequence of [any of claims 30-34] claim 30.

42. (Amended) A method of producing monensin comprising culturing the organism of claim 41 [and/or an organism produced by the method of claim 39 or claim 40].

POLYKETIDES AND THEIR SYNTHESIS

The present invention relates to processes and materials (including enzyme systems, nucleic acids, vectors and cultures) for preparing polyketides, particularly polyethers but including polyenes, macrolides and other polyketides by recombinant synthesis, and to the polyketides so produced, particularly novel polyketides. (N.B the term "polyketide" is being used in its conventional sense to include structures notionally derived by the reduction and/or other processing or modification of one or more Ketide units). Furthermore the invention provides the entire nucleic acid sequence of the biosynthetic gene cluster that governs the production of the ionophoric antibiotic polyether polyketide monensin in *Streptomyces cinnamonensis*, and the use of all or part of the cloned DNA first, in the specific detection of other polyether biosynthetic gene clusters; secondly in the engineering of mutant strains of *S. cinnamonensis* and of other actinomycetes which are suitable host strains for the high level production of novel recombinant polyketides; and thirdly in the provision of recombinant biosynthetic genes which lead to such novel polyketide products.

Polyketides are a large and structurally diverse

class of natural products that includes many compounds possessing antibiotic or other pharmacological properties, such as erythromycin, tetracyclines, rapamycin, avermectin, monensin, epothilones and FK506. In particular, polyketides are abundantly produced by *Streptomyces* and related actinomycete bacteria. They are synthesised by the repeated stepwise condensation of acylthioesters in a manner analogous to that of fatty acid biosynthesis. The greater structural diversity found among natural polyketides arises from the selection of (usually) acetate or propionate as "starter" or "extender" units; and from the differing degree of processing of the β-keto group observed after each condensation. Examples of processing steps include reduction to β-hydroxyacyl-, reduction followed by dehydration to 2-enoyl-, and complete reduction to the saturated acylthioester. The stereochemical outcome of these processing steps is also specified for each cycle of chain extension. In addition, the biosynthetic pathways to many polyketides involve additional enzyme-catalysed modifications which may include: methylation by O- and C-methyltransferases, hydroxylation by cytochrome P450 enzymes, other oxidation or reduction processes, and the biosynthesis and attachment of novel sugars and/or deoxy sugars.

The biosynthesis of polyketides is initiated by a group of chain-forming enzymes known as polyketide synthases. Two classes of polyketide synthase (PKS) have been described in actinomycetes. One class, named Type I PKSs, represented by the PKSs for the macrolides erythromycin, oleandomycin, avermectin and rapamycin, consists of a different set or "module" of enzymes for each cycle of polyketide chain extension. (For examples see Cortés, J. *et al.* Nature (1990) 348:176-178; Donadio, S. *et al.* Science (1991) 252:675-679; Swan, D.G. *et al.* Mol. Gen. Genet. (1994) 242:358-362; MacNeil, D.J. *et al.* Gene (1992) 115:119-125; Schwecke, T. *et al.* Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843.)

The term "extension module" as used herein refers to the set of contiguous domains, from a β-ketoacyl-ACP synthase ("KS") domain to the next acyl carrier protein ("ACP") domain, which accomplishes one cycle of polyketide chain extension. The term "loading module" is used to refer to any group of contiguous domains which accomplishes the loading of the starter unit onto the PKS and thus renders it available to the KS domain of the first extension module. The length of polyketide formed has been altered, in the case of erythromycin biosynthesis, by specific relocation using genetic engineering of the enzymatic domain of the erythromycin-

producing PKS that contains the chain releasing thioesterase/cyclase activity (Cortés J. *et al.* Science (1995) 268:1487-1489; Kao, C.M. *et al.* J. Am. Chem. Soc. (1995) 117:9105-9106).

In-frame deletion of the DNA encoding part of the ketoreductase domain in module 5 of the erythromycin-producing PKS (also known as 6-deoxyerythronolide B synthase, DEBS) has been shown to lead to the formation of erythromycin analogues 5,6-dideoxy-3-α-mycarosyl-5-oxoerythronolide B, 5,6-dideoxy-5-oxoerythronolide B and 5,6-dideoxy,6-β-epoxy-5-oxoerythronolide B (Donadio, S. *et al.* Science (1991) 252:675-679). Likewise, alteration of active site residues in the enoylreductase domain of module 4 in DEBS, by genetic engineering of the corresponding PKS-encoding DNA and its introduction into *Saccharopolyspora erythraea*, led to the production of 6,7-anhydroerythromycin C (Donadio, S. *et al.* Proc. Natl. Acad. Sci. USA (1993) 90:7119-7123).

International Patent Application number WO 93/13663 describes additional types of genetic manipulation of the DEBS genes that are capable of producing altered polyketides. However many such attempts are reported to have been unproductive (Hutchinson, C.R. and Fujii, I. Annu. Rev. Microbiol. (1995) 49:201-238, at p. 231). The complete DNA sequence of the genes from *Streptomyces*

hygroscopicus that encode the modular Type I PKS governing the biosynthesis of the macrocyclic immunosuppressant polyketide rapamycin has been disclosed (Schwecke, T. *et al.* (1995) Proc. Natl. Acad. Sci. USA 92:7839-7843). The DNA sequence is deposited in the EMBL/Genbank Database under the accession number X86780.

WO 98/01546 discloses that a PKS gene assembly (particularly of Type I) encodes a loading module which is followed by at least one extension module. The first open reading frame encodes the first multi-enzyme or cassette (DEBS1) which consists of three modules: the loading module (ery-load) and two extension modules (modules 1 and 2). The loading module comprises an acyltransferase and an acyl carrier protein. This may be contrasted with Figure 1 of WO 93/13663 (referred to above). This shows ORF1 as only two modules, the first of which is in fact both the loading module and the first extension module.

WO 98/01546 describes in general terms the production of a hybrid PKS gene assembly comprising a loading module and at least one extension module. It also describes (see also Marsden, A.F.A. *et al.* Science (1998) 279:199-202) construction of a hybrid PKS gene assembly by grafting the wide-specificity loading module for the avermectin-producing polyketide synthase onto the first

multi-enzyme component (DEBS1) for the erythromycin PKS in place of the normal loading module. Certain novel polyketides can be prepared using the hybrid PKS gene assembly, as described for example in WO 98/01571.

WO 98/01546 further describes the construction of a hybrid PKS gene assembly by grafting the loading module for the rapamycin-producing polyketide synthase onto the first multi-enzyme component (DEBS1) for the erythromycin PKS in place of the normal loading module. The loading module of the rapamycin PKS differs from the loading modules of DEBS and the avermectin PKS in that it comprises a CoA ligase domain, an enoylreductase ("ER") domain and an ACP, so that suitable organic acids including the natural starter unit 3,4-dihydroxycyclohexane carboxylic acid may be activated *in situ* on the PKS loading domain and, with or without reduction by the ER domain, transferred to the ACP for intramolecular loading of the KS of extension module 1 (Schwecke, T. *et al.* Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843). WO 98/51695 and WO 98/49315 describe additional types of genetic manipulation of the DEBS genes that are capable of producing altered polyketides.

The second class of PKS, named Type II PKSs, is represented by the synthases for aromatic compounds. Type II PKSs contain only a single set of enzymatic activities

for chain extension and these are re-used as appropriate in successive cycles (Bibb, M.J. *et al.* EMBO J. (1989) 8:2727-2736; Sherman, D.H. *et al.* EMBO J. (1989) 8:2717-2725; Fernandez-Moreno, M.A. *et al.* J. Biol. Chem. (1992) 267:19278-19290). The "extender" units for the Type II PKSs are usually acetate units, and the presence of specific cyclases dictates the preferred pathway for cyclisation of the completed chain into an aromatic product (Hutchinson, C.R. and Fujii, I. Ann. Rev. Microbiol. (1995) 49:201-238). Hybrid polyketides have been obtained by the introduction of cloned Type II PKS gene-containing DNA into another strain containing a different Type II PKS gene cluster, for example by introduction of DNA derived from the gene cluster for actinorhodin, a blue-pigmented polyketide from *Streptomyces coelicolor*, into an anthraquinone polyketide-producing strain of *Streptomyces galileus* (Bartel, P.L. *et al.* J. Bacteriol. (1990) 172:4816-4826).

The minimal number of domains required for polyketide chain extension on a Type II PKS when expressed in a *Streptomyces coelicolor* host cell (the "minimal PKS") has been defined for example in WO 95/08548 as containing the following three polypeptides which are products of the *actI* genes: firstly KS; secondly a polypeptide termed the CLF with end-to-end

amino acid sequence similarity to the KS but in which the essential active site residue of the KS, namely a cysteine residue, is substituted either by a glutamine residue or, in the case of the PKS for a spore pigment such as the *whiE* gene product (Davis, N.K. and Chater, K.F. Mol. Microbiol. (1990) 4:1679-1691) by a glutamic acid residue; and finally an ACP. The CLF has been stated (for example in WO 95/08548) to be a factor that determines the chain length of the polyketide chain that is produced by the minimal PKS. However it has been found (Shen, B. *et al.* J. Am. Chem. Soc. (1995) 117:6811-6821) that when the CLF for the octaketide actinorhodin is used to replace the CLF for the decaketide tetracenomycin in host cells of *Streptomyces glaucescens*, the polyketide product is not found to be altered from a decaketide to an octaketide, so the exact role of the CLF remains unclear. An alternative nomenclature has been proposed in which KS is designated KSα and CLF is designated KSβ, to reflect this lack of knowledge (Meurer, G. *et al.* Chemistry & Biology (1997) 4:433-443). The mechanism by which acetate starter units and acetate extender units are loaded onto the Type II PKS is not known, but it is speculated that the malonyl-CoA: ACP acyltransferase of the fatty acid synthase of the host cell can fulfil the same function for the Type II PKS (Revill, W.P. *et al.* J.

Bacteriol. (1995) 177:3946-3952).

WO 95/08548 describes the replacement of actinorhodin PKS genes by heterologous DNA from other Type II PKS gene clusters, to obtain hybrid polyketides. It also describes the construction of a strain of *Streptomyces coelicolor* which substantially lacks the native gene cluster for actinorhodin, and the use in that strain of a plasmid vector pRM5 derived from the low-copy number vector SCP2* isolated from *Streptomyces coelicolor* (Bibb, M.J. and Hopwood, D.A. J. Gen. Microbiol. (1981) 126:427-442) and in which heterologous PKS-encoding DNA may be expressed under the control of the divergent *actI/actIII* promoter region of the actinorhodin gene cluster (Fernandez-Moreno, M.A. *et al.* J. Biol. Chem. (1992) 267:19278-19290). The plasmid pRM5 also contains DNA from the actinorhodin biosynthetic gene cluster encoding the gene for a specific activator protein, ActII-orf4. The ActII-orf4 protein is required for transcription of the genes placed under the control of the *actI/actIII* bidirectional promoter and activates gene expression during the transition from growth to stationary phase in the vegetative mycelium (Hallam, S.E. *et al.* Gene (1988) 74:305-320).

Type II clusters in *Streptomyces* are known to be activated by pathway-specific activator genes (Narva,

K.E. and Feitelson, J.S. J. Bacteriol. (1990) 172:326-333; Stutzman-Engwall, K.J. *et al.* J. Bacteriol. (1992) 174:144-154; Fernandez-Moreno, M.A. *et al.* Cell (1991) 66:769-780; Takano, E. *et al.* Mol. Microbiol. (1992) 6:2797-2804; Gramajo, H.C. *et al.* Mol. Microbiol. (1993) 7:837-845). The DnrI gene product complements a mutation in the *actII*-orf4 gene of *S. coelicolor*, implying that DnrI and ActII-orf4 proteins act on similar targets. A gene (*srmR*) has been described (EP 0 524 832 A2) that is located near the Type I PKS gene cluster for the macrolide polyketide spiramycin. This gene specifically activates the production of the macrolide antibiotic spiramycin, but no other examples have been found of such a gene. Also, no homologues of the ActII-orf4/DnrI/RedD family of activators have been described that act on Type I PKS genes. WO 98/01546 describes the use of the ActII-orf4 family of activators in conjunction with their cognate promoters (e.g actII-orf4 with the *actI* promoter) in a heterologous actinomycete to obtain high level expression of recombinant Type I polyketide synthase genes.

Although large numbers of therapeutically important polyketides have been identified, there remains a need to obtain novel polyketides that have enhanced properties or possess completely novel bioactivity. The complex

polyketides produced by Type I PKSs are particularly valuable, in that they include compounds with known utility as anthelminthics, insecticides, immunosuppressants, antifungal agents or antibacterial agents. Because of their structural complexity, such novel polyketides are not readily obtainable by total chemical synthesis, nor by chemical modifications of known polyketides.

There is also a need to develop reliable and specific ways of deploying individual genes and portions of genes in practice so that all, or a large fraction, of hybrid PKS genes that are constructed, are viable and produce the desired polyketide product. This includes the development of advantageous host strains for expression of such genes. For example many polyketides are rendered bioactive by the action of further enzymes other than the polyketide synthase, and host strains that contain and are able to express the genes for such enzymes are particularly convenient for the efficient synthesis of the bioactive material. In those cases where the construction of a known or a novel polyketide requires specialised precursors, host strains containing and able to express the genes for key enzymes that enhance the production of such specialised precursors are equally valuable and desirable. There is also a need to develop

rational methods of increasing the expression level of all the genes required for production of a specific polyketide. Clearly also a host cell which is advantageous for the above reasons, and/or because of other favourable characteristics including but not limited to its speed of growth, excellent handling characteristics in fermentation, and ease of transformation with DNA by various techniques, can be made even more favourable by the cloning into that cell of such auxiliary genes for polyketide modification, or gene activation, or post-translational modification, or precursor supply.

The DNA sequences have been disclosed for several Type I PKS gene clusters that govern the production of 16-membered macrolide polyketides, including the tylosin PKS from *Streptomyces fradiae* (application EP 0 791 655 A2), the niddamycin PKS from *Streptomyces caelestis* (Kavakas, S.J. *et al.* J. Bacteriol. (1997) 179:7515-7522) and the spiramycin PKS from *Streptomyces ambofaciens* (application EP 0791 655 A2). DNA sequences have also been disclosed for Type I PKS gene clusters that govern the production of further complex polyketides, for example rifamycin from *Amycolatopsis mediterranei* (WO 98/07868), and soraphen from *Sorangium cellulosum* (US

5716849), but so far no DNA sequence has been disclosed for one of the most widespread and important classes of complex polyketides, the polyethers.

Polyethers form an important group of complex polyketide antibiotics (Westley, J.W. in "Antibiotics IV. Biosynthesis" (Corcoran, J.W. Ed.), Springer-Verlag, New York (1981) p. 41-73). They are polyoxygenated carboxylic acids which act as selective ionophores transporting cations across the cell membrane of target cells and thereby causing depolarisation and cell death. Certain polyethers including monensin, lasalocid and tetronasin are in widespread use in animal husbandry as coccidiostats (principally targetted against *Eimeria spp.*) and as growth promoters. Polyethers have also been reported to be active *in vitro* and *in vivo* against the malarial parasite *Plasmodium falciparum* (Gumila, C. *et al.* Antimicrobial Agents and Chemotherapy (1997) 41: 523-529).

Polyethers contain multiple asymmetric centres and are characterised by the presence of tetrahydrofuran and tetrahydropyran rings, producing a characteristic shape which is non-polar on its outer surface and therefore well adapted for transport of material across bacterial membranes; and provides on its inner surface polar coordinating ligands for a centrally-bound metal ion. In

addition to tetrahydrofuran and tetrahydropyran rings, other groups which are often present include spiroketal, dispiroketal, and substituted benzoic acid moieties and occasionally other groups for example a tetronic acid or a 6-membered carbocyclic ring

Monensins A and B are produced by the actinomycete *Streptomyces cinnamonensis*. Their structures are shown in Figure 1. Monensin B differs from monensin A only in the presence of a methyl sidechain at C-16 rather than an ethyl sidechain. Monensin selectively binds and transports sodium ions. In addition to its antibacterial and antifungal properties monensin has some activity against protozoal parasites such as the malarial parasite *Plasmodium falciparum*. Although the structures of polyethers differ significantly from those of other complex polyketides such as the polyhydroxylated and polyene macrolides, their biosynthesis appears to take place by a metabolic pathway which has many common elements. Thus experiments using carbon 14-labelled precursors have shown that monensin A is synthesised from five acetate, one butyrate and seven propionate units (Day, L.E. *et al.* Antimicrob. Agents Chemother. (1973) 4:410-414). Similarly experiments using precursors doubly-labelled with carbon-13 and oxygen-18 have shown that oxygens (O)1, (O)3, (O)4, (O)5, (O)6 and (O)10 of

monensin arise from the carboxylate oxygens of either propionate or acetate, while growth in the presence of oxygen-18 oxygen gas demonstrated that the three remaining ether oxygens (O)7, (O)8 and (O)9 are derived from molecular oxygen (Cane, D.E. *et al.*, J. Am. Chem. Soc. (1981) 103:5962-5965; Cane, D.E. *et al.* J. Am. Chem. Soc. (1982) 104:7274 - 7281; Ajaz, A.A. and Robinson, J.A. J. Chem. Soc. Chem. Commun. (1983) 12:679-680). These findings have been rationalised by proposing that the biosynthesis of monensin proceeds via an acyclic triene intermediate (1) in which the geometry of all three carbon-carbon double bonds is E (entgegen) rather than Z (zusammen). The triene is then proposed to be subject to epoxidation to a tri-epoxide (2) and then ring opening is proposed to occur with concomitant sequential formation of the five ether rings as shown in Figure 2A. Such a biosynthetic pathway, first mooted by Westley in 1974 (Westley J.W. *et al.*, J. Antibiot. (1974) 27:597-604) accounts for the observed stereochemistry at the multiple asymmetric centres in monensin, (Cane, D.E. *et al.* J. Am. Chem. Soc. (1982) 104:7274-7281; Sood, G.R. *et al.* J. Chem. Soc. Chem. Commun. (1984) 21:1421-1424) and analogous schemes can be used to account for the biosynthesis of other known polyethers. such as lasalocid A (Hutchinson C.R. *et al.*, J. Am. Chem. Soc. (1981)

103:5953-5956), tetronasin (ICI 139603) (Demetriadou, A.K. *et al.* J. Chem. Soc. Chem. Commun. (1985) 7:408-410) and narasin (Spavold, Z. *et al.* Tetrahedron Letters (1986) 27:3299-3302). The hydroxylation at C-26 and the introduction of an O-methyl group on oxygen 3-are proposed to occur as late steps in the biosynthesis, after formation of the polyether structure.

Unfortunately key aspects of the biosynthetic scheme shown in Figure 2A have so far eluded experimental confirmation. No biosynthetic intermediates have been isolated from mutants of *S. cinnamonensis* that are blocked in early stages of monensin production. 26-deoxymonensin A has been isolated from a *S. cinnamonensis* mutant partially blocked in monensin production (Ashworth, D.M. *et al.* J. Antibiot. (1989) 42:1088-1099) and 3-O-demethylmonensins A and B have been recovered as minor components from the fermentation broth of a monensin-producing strain (Pospisil, S. *et al.* J. Antibiot. (1987) 40:555-557). When fed to cells of *S. cinnamonensis* in radio-labelled form, neither 26-deoxymonensin A, nor 3-O-demethylmonensin A, nor 3-O-demethyl, 26-deoxymonensin A were significantly incorporated into monensin A (Ashworth, D.M. *et al.* J. Antibiot. (1989) 42:1088-1099), either because they are actively excluded or because these modifications in fact

occur earlier in the biosynthetic pathway so that these metabolites are shunt products not readily converted into the final antibiotic by the respective hydroxylase or methyltransferase. Similarly, the putative all (E)-triene precursor (1) has been synthesised and shown not to become incorporated into monensin when fed to growing cells of *S. cinnamonensis* (Holmes, D.S. *et al.* Helv. Chim. Acta (1990) 73:239-259). An alternative pathway has been proposed, as shown in Fig 2B, based on the transition-metal-mediated oxidation of 1,5-dienes (Walba, D.M. and Edwards, P.D. Tetrahedron Lett. (1980) 21:3531-3534). The triene intermediate (4) would different from that of Figure 2A (1) only in that each carbon-carbon double bond would have the (Z)-configuration (Townsend, C.A. and Basak, A. Tetrahedron (1991) 47:2591-2602) and not the (E)- configuration.

The genetic basis of secondary metabolite biosynthesis essentially exists in the genes which code for the individual biosynthetic enzymes and in the regulatory elements which control the expression of the biosynthetic genes. The genes encoding biosynthesis of polyketides in actinomycetes have hitherto been found as clusters of adjacent genes, ranging in size from 20 kilobasepairs (kbp) to over 100 kbp. The clusters often contain specific regulatory genes and genes

conferring resistance of the producing strain to its own antibiotic.

In various of its aspects the invention provides the following:-

(1) a DNA sequence encoding at least one peptide necessary for the biosynthesis of monensin, preferably comprising one or more of the following genes: *mon BI, mon BII, mon CI, mon CII, mon H, mon RI, mon RII, mon T, mon AIX* and *mon AX* as depicted in the appended sequence data or an allele or mutation thereof;

(2) a DNA sequence according to the first aspect comprising all of the genes listed therein or an allele or mutation thereof;

(3) a DNA sequence according to the first aspect comprising the complete monensin gene cluster;

(4) a DNA sequence coding for one or more of the peptides set out below, said peptide having the amino acid sequence as set out in the appended sequence data or being a variant thereof having the specified activity:

peptide	activity
mon CII	epoxyhydrolase/cyclase
mon E	S-adenosylmethionine-dependent methyltransferase
mon T	monensin resistance gene
mon RII	repressor protein
mon AIX	thioesterase

mon AI	polyketide synthase multienzyme
mon AII	polyketide synthase multienzyme
mon AIII	polyketide synthase multienzyme
mon AIV	polyketide synthase multienzyme
mon AVI	polyketide synthase multienzyme
mon AVII	polyketide synthase multienzyme
mon AVIII	polyketide synthase multienzyme
mon H	regulatory protein
mon CI	flavin-dependent epoxidase
mon BII	carbon-carbon double bond isomerase
mon BI	carbon-carbon double bond isomerase
mon D	cytochrome P450 hydroxylase
mon RI	activator protein
mon AX	thioesterase

(5) a recombinant cloning or expression vector comprising a DNA sequence according to any of aspects 1-4;

(6) a transformant host cell which has been transformed to contain a DNA sequence according to any of aspects 1-4 and is capable of expressing a corresponding peptide;

(7) a hybridization probe comprising a polynucleotide which binds specifically to a region of the monensin gene cluster selected from *mon BI, mon BII, mon CI, mon CII, mon H, mon RI, mon RII, mon T, mon AIX* and *mon AX;*

(8) use of a probe according to aspect (7) in a method of detecting the presence of a gene cluster which governs the synthesis of a polyether, and optionally isolating a gene cluster detected thereby;

(9) Use of a probe comprising a polynucleotide which binds specifically to a gene responsible for levels of activity of the monensin gene cluster, preferably a regulatory gene, resistance gene or thioesterase gene, more preferably the regulatory gene *mon RI*, in a method of detecting an analogous gene in a gene cluster of another polyketide, preferably a polyether, and optionally manipulating the gene detected thereby to alter the level of expression of said other polyketide;

(10) a host cell, preferably *Streptomyces cinnamonensis*, containing a heterologous gene under the control of the *mon RI* gene and a monensin promoter;

(11) use of a portion of the monensin gene cluster having chain terminating activity, preferably comprising at least one of *mon AIX* and *mon AX* or a mutant or allele thereof having chain terminating activity, to effect chain release of a peptide other than one required for monensin biosynthesis;

(12) use of a portion of the monensin gene cluster having carbon-carbon double bond isomerase activity, preferably comprising at least one of *mon BI* and *mon BII*

or a mutant or allele thereof having isomerase activity to provide a desired stereochemical outcome in the synthesis of a polyketide other than monensin;

(13) a polypeptide encoded by a portion of the monensin gene cluster, preferably comprising at least one of *mon BI* and *mon BII* or a mutant or allele thereof, having carbon-carbon double bond isomerase activity;

(14) an epoxidase enzyme encoded by *mon CI* or a derivative or variant thereof having epoxidase activity;

(15) a cyclase enzyme encoded by *mon CII* or a derivative or variant thereof having cyclase activity.

Some embodiments of the invention will now be described by way of example with reference to the accompanying drawings in which:

Fig 1 shows the structure of monensins A and B;

Fig 2 illustrates proposed biosynthetic pathways;

Fig 3 illustrates the proposed organization of the monensin polyketide synthase (PKS) enzyme complex; and

Fig 4 illustrates the proposed organization of the monensin biosynthetic gene cluster.

The overall gene organization of the monensin biosynthetic gene cluster, as shown in Fig 4, is similar to that previously found for many macrolide biosynthetic gene clusters, which have one or more open reading frames (ORFs) encoding large multifunctional PKSs flanked by

other genes which encode functions required for the biosynthesis of the antibiotic. In the case of monensin, there is an unusually high number of distinct ORFs encoding PKS multi-enzymes (eight in total, labelled *monAI* to *monAVIII*) but there is again a separate module of enzymes for each cycle of polyketide chain extension, exactly as found for modular PKSs for macrolide biosynthesis (see Fig 3). Thus there are 12 condensations predicted to be required for the production of the carbon skeleton of monensin, and in agreement with this there are found to be 12 extension modules of PKS enzymes distributed among the 8 PKS ORFs. However, as mentioned in detail below, the other genes in the monensin cluster include genes which have not previously been found in any other gene cluster for the biosynthesis of a complex polyketide, and which are not significantly similar to any genes in published sequence databases. The cloned DNA for these genes is useful to allow the diagnosis that a polyketide biosynthetic gene cluster in any actinomycete, uncovered previously by conventional hybridization against a PKS gene probe from (say) the DEBS or some other characterised PKS gene cluster, is one that governs the synthesis of a polyether; and these genes are also valuable either singly or in combination as specific hybridization probes for the specific detection and

isolation of additional polyether biosynthetic gene clusters. Examples of these previously-unknown genes are the genes *monBI*, *monBII*, *monCI* and *monCII*. In addition the regulatory genes *monH* *monRI*, and *monRII* and the resistance gene *monT* and the thioesterase genes *monAIX* and *monAX* are all useful for the detection of analogous genes in other polyether clusters which are required for the rational manipulation of such genes in order to increase levels of the specific product.

The cloned and sequenced cluster of genes for monensin biosynthesis is useful secondly in the engineering of mutant strains of *S. cinnamonensis* and of other actinomycetes which are suitable strains for the high level production of either natural or novel recombinant polyketides. The sequence of the monensin cluster disclosed here shows the surprising fact, that the gene cluster contains a gene *monRI* whose gene product has an amino acid sequence highly similar to that of *actII*-orf4, the pathway-specific activator gene which activates the *actI* and other promoters of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor*. The recognition of this aspect of the natural regulation of a Type I PKS cluster is important and valuable because first, it is possible to increase the yield of monensin by increasing the level of the activator MonRI, either by

placing the gene *monRI* under the control of a powerful promoter or arranging for the presence within the cells of one or more additional copies of the *monRI* gene (as exemplified below); secondly, it will be possible to use the *monRI* gene as a specific hybridisation probe to locate similar genes in other complex PKS gene clusters, especially other polyether PKS gene clusters but also polyene and macrolide gene clusters and all other Type I modular PKS gene clusters; even in cases where (as for rapamycin and erythromycin) no such gene has been previously found within the currently accepted physical limits of the relevant biosynthetic gene cluster. In such cases the *monRI* gene probe might be expected to uncover the activator even if it resides on the chromosome at some distance from the main body of the gene cluster; and simple experiments would then show whether the activator(s) so uncovered are involved in regulation of the biosynthesis of those particular metabolites; thirdly, increasing the copy number of the *monRI* gene or of any of the activator genes uncovered will tend to increase the yield of a heterologous polyketide by "crosstalk" where the activator mimics the presence of the normal activator for the transcription of the genes for that heterologous polyketide synthase. It is clear from recently published work (Wietzorrek, A. and Bibb, M. Mol. Microbiol. (1997)

25:1181-1184) that the ActII-orf4 family of activators exert their effects by binding to promoter regions within the target gene cluster, so it will be possible to use the *monRI* gene together with monensin promoter regions to drive the high-level transcription and translation of heterologous genes in *Streptomyces cinnamonensis*, and perhaps in other host strains too; such genes need not be PKS genes or even involved in polyketide biosynthesis. Monensin promoter regions are found at the 5' end of genes or groups of genes in the cluster and their location is clear from the sequence analysis disclosed here. Thus a useful vector would provide the monensin promoter and the ribosome binding site and continue up to the start of the open reading frame, after which the monensin ORF naturally found there would be replaced by the heterologous gene. The relative strength of the monensin promoters can be readily determined using any one of a number of known promoter probes, i.e. genes whose expression gives rise to readily measurable and quantifiable effects, such as Green Fluorescent Protein (GFP); or beta-galactosidase in the presence of a chromogenic substrate. It should be possible to mutate randomly the small region of the monensin promoters especially likely to interact with the MonRI activator (identified by the presence of tandem heptanucleotide repeats with a common consensus sequence

between the various monensin promoters) (Wietzorrek, A. and Bibb, M. Mol. Microbiol. (1997) 25:1181-1184), and to determine the optimal DNA sequence for the maximal activation effect using either *S. cinnamonensis* (preferably - in case there are other unknown factors that make the activation function better in this strain than in other heterologous systems), or even in another host actinomycete strain. If the natural monensin promoters were mutated to have this optimal recognition sequence, then this would further increase the production of monensin. By extension, the use of this modified monensin promoter in conjunction with the *monRI* gene in heterologous systems could form the basis of further improvements in expression of polyketide synthases or other genes, either by appropriate chromosomal alterations to introduce the altered promoter and also the *monRI* gene; or by provision of vectors containing these optimised signals linked to specific genes and housed in suitable host cells.

The sequencing of the monensin cluster has uncovered another strategy for gene regulation in such Type I clusters. The previously-sequenced genes for the rapamycin biosynthetic pathway in *Streptomyces hygroscopicus* included a gene of unknown function (*rapH*). A closely similar gene has now been found in the monensin

biosynthetic gene cluster (*monH*), and it is clear from this recurrence (and the comparison of the sequences with those of database proteins) that this gene is potentially an important DNA-binding sensor gene which acts to regulate the transcription of the cluster in concert with other regulatory signals. Simple experimentation is needed in order to define whether the gene is an activator, in which case putting in another copy or increasing its transcription will have the potential to increase polyketide biosynthesis; or alternatively the *rapH* gene product may be a negative regulator, whereupon deletion of this gene may release the biosynthetic pathway from this inhibitory effect and increase yields.

There is a continuing need to develop new methods of high-level production of bioactive metabolites and other valuable gene products in actinomycetes. *Streptomyces cinnamonensis* is a recognised and very valuable industrial strain for the production of very high levels of monensin, it is readily transformable with DNA by standard methods of conjugation or of protoplast transformation, it is a host for numerous known broad range plasmids including well-known expression plasmids of both high- and low-copy number, it also grows quickly relative to other actinomycete strains (for example about three times faster than wild type *Saccharopolyspora erythraea* the

erythromycin producer, under comparable conditions) and sporulates relatively easily. Heterologous polyketides can be expressed in *Streptomyces cinnamonensis* using for example the low-copy number plasmid pCJR24 (which has no origin of replication active in actinomycetes so is maintained by integration into the chromosome) (Rowe, C. *et al.* Gene (1998) 216:215-223) or the related plasmid pCJR29 in which the polyketide synthase gene(s) are placed under the control of the *actI* promoter which is activated by the ActII-orf4 activator; or alternatively the *monAI* promoter can be substituted together with the MonRI activator; or some other pairing of activator and cognate promoter chosen from either a Type II or a Type I polyketide synthase gene cluster. As an example, the wild type strain of *Streptomyces cinnamonensis* has been used to express the plasmid pCJR29 (Rowe, C. *et al.* Gene (1998) 216:215-223) containing as insert the three ORFs for the PKS governing the production of 6-deoxyerythronolide B, the macrolide precursor of erythromycin A in *Saccharopolyspora erythraea*, these genes being placed under the control of the pathway-specific *actI* promoter from *Streptomyces coelicolor* together with its cognate activator gene *actII*-orf4. The transformed strain when cultivated in a suitable liquid medium produced 6-deoxyerythronolide B in good yield.

It is well known to the person skilled in the art that it is possible to use standard vectors unable to replicate in actinomycetes to introduce DNA into a *Streptomyces* cell, such DNA comprising two portions of contiguous DNA which are each identical to one of two portions of the cell's chromosome that are spaced up to 100 kbp apart; and that through recombination between the incoming DNA and the chromosome occurring in both portions of DNA the net result is that the chromosomal sequence is replaced by the defective sequence originally that of the incoming DNA. Such a procedure has been applied to the monensin-producing strain of *S. cinnamonensis* as described in detail below, and a strain of *S. cinnamonensis* has been obtained that carries a specific deletion in the monensin cluster and which is unable to produce the antibiotic. The use of such a strain facilitates the production of heterologous polyketides by removal of the background of monensin production.

The multiple uses of portions of the cloned and sequenced DNA from the monensin cluster will readily occur to the person skilled in the art. A surprising feature of the PKS of the monensin cluster is an unusual mechanism of polyketide chain initiation. We have found that the monensin PKS loading module has three domains, which from the amino-terminus of the protein are: a KSq domain, an

acyltransferase domain and an ACP domain. We have uncovered this organisation in the PKS for the 14-membered macrolide oleandomycin as well as in the monensin PKS, an organisation of the loading module previously only found for the 16-membered macrolides and in which the KSq domain (which looks like a ketosynthase or condensation domain except that the active site cysteine residue is substituted by a glutamine for which the single letter notation is Q) had been previously speculated to have no function. It was realised that the acyltransferase of the loading module actually has malonyl-CoA and not acetyl-CoA as a substrate and that KSq is an active decarboxylase. It appears that a better discrimination can be achieved in the selection of the smaller acetate unit over propionate if the choice is made initially between methylmalonyl- and malonyl-CoA.

An unprecedented feature of the monensin PKS genes is that no integral chain-terminating domain is present as a C-terminal appendage of the PKS extension module that catalyzes the twelfth and final chain extension. Because the product of the monensin PKS ~~is~~ a carboxylic acid, it would have been firmly predicted that chain release would have been catalyzed by such a C-terminal domain containing a "thioesterase" activity. Previously sequenced PKS gene sets have been of two sorts: first, those macrolide PKSs

typified by erythromycin, spiramycin, tylosin, niddamycin which have a readily recognisable C-terminal "thioesterase" domain, which in these enzymes functions as a specific cyclase rather than releasing the polyketide product as a free carboxylic acid; secondly, those macrolide PKSs typified by rapamycin, FK506, and rifamycin, where there is an alternative and recognised mode of chain termination by transfer of the polyketide chain to an acceptor moiety, catalyzed by a specific enzyme (eg pipecolate incorporating enzyme for rapamycin (Schwecke T. *et al.* Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843) and FK506 (Mothamedi H. and Shafiee A, Eur. J. Biochemistry (1998) 256:528-534); arylamine synthetase for rifamycin (August P.R. *et al.* Chemistry & Biology (1998) 5:69-79).

The monensin PKS surprisingly falls into neither category, and therefore seems to be the first example of a novel mode of chain termination. It is novel and noteworthy in this connection that the monensin PKS gene cluster contains two small genes that encode discrete, monofunctional thioesterase enzymes. Although many PKS gene clusters have been previously shown to contain one such discrete thioesterase, none have been shown to have two. The role of such thioesterases is not known, although in the case of methymycin/pikromycin PKS, which has been

reported to be responsible for the biosynthesis of both the 12-membered macrolide methymycin and the 14-membered macrolide pikromycin (Xue Y.Q. Proc. Natl. Acad. Sci. USA (1998) 95:12111-12116) the disruption of this thioesterase reportedly caused a ten-fold drop in the amount of both macrolides produced. A similar finding has been reported for the discrete thioesterase of the tylosin PKS gene cluster (Cundliffe E. *et al.* Chemistry & Biology in press). Additional copies of such thioesterases may therefore accelerate the production of specific polyketide, but this has not yet been demonstrated. However, the presence of the discrete thioesterase is not completely essential for polyketide production.

It is highly desirable to have a broadly effective method of catalysing the release of polyketide gene products from a PKS as the free acid. The well-studied integral thioesterase domain in the erythromycin PKS thioesterase has a broad specificity in cyclization to form a lactone (assuming that a hydroxy group is present in the growing polyketide chain at an appropriate position), but hydrolysis to form the free acid is very slow. The recognition of the unusual arrangement of the monensin PKS means that it is now possible to harness either the entire PKS module that catalyses the twelfth and final extension cycle in monensin biosynthesis, or the

C-terminal portion of it, and graft it onto a different polyketide synthase by genetic engineering, so as to allow the release mechanism characteristic of monensin to operate in a different context. The use of this portion only of the monensin PKS suffices to allow the novel mechanism of chain release to operate successfully. The speed of the polyketide chain hydrolysis in a given case can depend on the additional presence of one or both of the discrete thioesterase genes (*monAIX* and *monAX*) from the monensin gene cluster. The use of this novel method of chain termination represents a valuable way of generating a large number of novel engineered polyketides that are currently inaccessible, and ensuring that the products have a specified chain length.

The genes *monBI* and *monBII* appear to encode very similar enzymes with significant amino acid sequence similarity to authentic ketosteroid isomerases which are known to catalyse the migration of an activated carbon-carbon double bond. The conservation of active site residues makes it very likely that these *mon* genes govern a reaction involving activated double bonds in the biosynthetic pathway to monensin and this surprising observation can be accommodated if the initial product of the polyketide chain growth on the monensin PKS is a linear precursor in which the double bonds were initially

formed with a conventional *trans* or E (entgegen) geometry; but before the polyketide chain was extended by insertion of the next unit the *monBI* and/or the *monBII* gene product(s) catalyse the specific rearrangement of the newly-created double bond into the *cis* or Z (zusammen) geometry. This new view of the monensin biosynthetic pathway allows the deduction that the *monBI* and *monBII* genes, perhaps in combination with specific portions of the monensin modules where they normally exert their effects (namely modules 3, 5 and 7) might be used in order to achieve the extremely desirable targetted biosynthesis of novel polyketides containing double bonds with Z geometry at specified point(s) along the chain. Thus for example it should be possible to provide for the direct biosynthesis of C22-C23 *cis* or Z double bond in avermectins, thus avoiding tedious and expensive chemical conversion of an initial fermentation product into this important anthelminthic. Only limited experimentation is needed to see whether the *monBI* and/or *monBII* gene products are sufficient or whether the *mon* PKS at modules 3, 5 and 7 forms part of the specific docking site(s) for the isomerases and therefore must also be used in the creation of the hybrid PKS that will insert the *cis* or Z double bond at the desired position. The substrate specificity of the isomerases need not be limited to 2,3-

unsaturated thioesters. The purified enzymes could also be used to effect such isomerisations *in vitro*, depending on the position of the equilibrium or whether further enzymes are used to achieve the further transformation of the product as it is formed (*vide infra*).

The product of the *monCI* gene is a novel oxidative enzyme with some sequence similarity to authentic examples of such enzymes in the databases; and with a clearly definable role in the monensin biosynthetic pathway, the epoxidation of the double bonds at three separate positions in the initially-formed acyclic intermediate in monensin biosynthesis. This epoxidase could therefore be used in conjunction with *monBI/monBII* gene products to effect oxidative reactions on suitable substrates *in vitro* and *in vivo*. Similarly the *monCII* gene product is a putative cyclase that opens the epoxides and causes the formation of ether rings in monensin.

Any or all of the *monBI, monBII, monCI or monCII* genes may be introduced into a heterologous strain containing the gene cluster for another polyether, in order to divert the biosynthetic pathway and produce a polyketide of altered structure. In these experiments the analogues of these *monB* genes could either be present or (once located and characterised using the *mon* genes as probes) they may be deleted prior to the introduction of

the *monB* and *monC* genes into that strain. The converse experiment in which analogues of the *monB* and *monC* genes from other strains are introduced into *S. cinnamonensis* likewise has the potential to produce novel oxidised polyketides. Also, the *monB* and *monC* genes or their analogues may be introduced into a strain that normally produces a macrolide or a polyene or some other complex polyketide and expressed there, when they may effect the diversion of the growing polyketide chain on a heterologous modular PKS towards a new product, which may or may not have the structure of a polyether.

The availability of the monensin gene sequence allows the institution of domain swaps to alter the acyltransferase (AT) specificity of a given module, for example the ethylmalonyl-CoA specific extender found in one of the modules of the monensin PKS can be used to replace one of the other ATs to generate an ethyl side branch at that position in the chain, or the AT can be used to substitute in any other (e.g. macrolide) PKS, as described in WO 98/01571 and WO 98/01546. Similarly the alteration of the level of reduction in a module, by manipulation of the reductive enzymes, can be applied to the monensin genes and here it will produce, depending on which module is affected, either an altered monensin, or a

species which is only partly cyclised, or a polyether with an altered pattern of cyclisation, or even a linear polyketide.

In general the targetted alteration of the pattern of substitution of sidechains or reduction level along the polyketide chain produced by the monensin PKS will, like the disruption or deletion of the oxidative enzymes mentioned above, lead to non-polyether polyketide products. It should be possible, by introduction of the DEBS thioesterase at the C-terminus of one of the later modules of the monensin PKS, together with an appropriately placed hydroxy group earlier in the chain, to produce novel macrolide products from this polyether PKS system, or alternatively novel polyenes of defined chain length and chosen ring size.

Example 1

Cloning of the monensin A biosynthetic gene cluster using DNA probes derived from the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*

A genomic library of the monensin A producing strain *Streptomyces cinnamonensis* ATCC 15413 was constructed using methods well-known in the art, namely, the production of high molecular weight genomic DNA, followed by the partial cleavage of this DNA using the frequent-cutting restriction enzyme *Sau*3A, fractionation of the fragments on a sucrose gradient and selection of fragments of average size 35-40 kbp, and the cloning of these fragments into the cosmid vector pWE15 (Evans, G.A. *et al.* Gene (1989) 79:9-20) which had been previously digested with *Bam*HI and treated with shrimp alkaline phosphatase. The library was packaged and transfected into *Escherichia coli* XL-1 Blue MR cells. The library was plated out on 2xTY agar medium (10 g tryptone, 10 g yeast extract, 5 g NaCl, 15 g bactoagar per litre containing ampicillin 50 μg/ml) for cosmid selection and the colonies were allowed to grow overnight. The library was then screened by hybridisation using as a probe DNA encoding the ketosynthase domain of module 1 of the erythromycin-producing PKS (6-deoxyerythronolide B synthase, DEBS) of *Saccharopolyspora erythraea*. The colonies giving a

positive hybridisation signal in the hybridisation were selected and the cosmid DNA from each colony was purified and mapped by restriction digestion. The presence of the target biosynthetic genes on a cosmid was verified by sequencing of the ends of the cosmid inserts using the commercially available T3 and T7 primers which hybridise specifically to the respective ends of each cosmid insert (Evans, G.A. *et al.* Gene (1989) 79:9-20).

Example 2

Sequencing of the biosynthetic gene cluster for monensin A from *Streptomyces cinnamonensis*

Three cosmids obtained by screening of the genomic library of *S. cinnamonensis* were used to obtain the entire DNA sequence of the monensin biosynthetic gene cluster. These cosmids, MO.CN02, MO.CN11 and MO.CN33 between them contain the entire DNA sequence of the cluster and the adjacent regions of the chromosome. They have been deposited in NCIMB, 23 St Machair Drive, Aberdeen AB24 3RY, UK, under the NCIMB accession numbers 40956 (MO-CN11); 40957 (MO-CN33) and 40958 (MO-CN02) respectively.

The DNA of each cosmid was separately subjected to partial digestion with *Sau*3A and fragments of approximately 1.5-2.0 kbp were separated by agarose gel electrophoresis. The fragments were then ligated into the

plasmid vector pUC18 (Messing, 1982), previously digested with *Bam*HI and treated with shrimp alkaline phosphatase. The library was transformed into *E. coli* strain XL1-Blue MR and plated on 2xTY agar medium containing ampicillin (100 µg/ml) to select for plasmid-containing cells. Plasmid DNA was purified from individual colonies and sequenced using the Sanger dye-terminator procedure on an ABI 377 automated sequencer (Sanger, F. Science (1981) 214:1205-1210). The sequence data obtained from single random subclones of a cosmid was assembled into a single continuous sequence and edited using GAP4.1 program of the STADEN gene analysis package (Staden, R. Molecular Biotechnology (1996) 5:233-241).

The sequence is set out in the appended sequence listing.

Tables I and II contain data about individual genes and gene products.

Example 3

Inactivation of the monensin A biosynthetic gene cluster

A chromosomal gene disruption experiment was used to verify the identity of the cloned polyketide synthase gene cluster. Plasmid pMOB6314 is a pUC18 sequencing subclone of the presumed monensin A biosynthetic gene cluster prepared as described in Example 1, whose inserted DNA comprises the DNA sequence from nucleotide 9763 to

nucleotide 10108 in SEQ ID 1, and which therefore contains a region of DNA wholly internal to *orfE*, a putative 3-O-methyltransferase. A *Hin*dIII fragment containing the thiostrepton resistance gene *tsr* from plasmid pIJ702 (Katz, E. *et al.* J. Gen. Microbiol. (1983) 129:2703-2714) was cloned into the *Hin*dIII site of plasmid pMOB6314 and the ligation mixture was used to transform *E. coli* cells. Transformants bearing the required plasmid pMOΔE01 were identified by isolation of plasmid DNA and analysis by restriction digestion. pMOΔE01. Plasmid pMOΔE01 was used to transform protoplasts of *Streptomyces cinnamonensis* as described by (Hopwood D.A. *et al.* (1985)). Since plasmid pMOΔE01 lacks an origin of replication that is active in *Streptomyces*, growth in the presence of thiostrepton (25 µg/ml) in the regeneration medium led to the isolation of stable integrants. Isolated putative integrants were tested for the presence of integrated pMOΔE01 sequences by Southern hybridisation. A clone of *Streptomyces cinnamonensis* identified by its restriction pattern in Southern hybridisation as bearing pMOΔE01 integrated in the region of *monE* of the monensin A biosynthetic gene cluster was designated *S. cinnamonensis* MO-DD01.

Detection of production of the monensin A related metabolites produced by *S. cinnamonensis* MO-DD01 was performed by GC-MS analysis of methanol extracts of the

entire broth harvested in 72 hours of growth of the strain. No significant amounts of monensin A-related metabolite production were detectable.

Example 4

Overproduction of erythromycin aglycone in *Streptomyces cinnamonensis*

S. cinnamonensis is a suitable system for overproduction not just of monensin A but also of other polyketide metabolites. Established techniques of genetic transformation allow fast introduction of foreign polyketide producing genes sets into this host. Fast growth of *S. cinnamonensis* in liquid culture and optimal precursor supply favour high yield of polyketide metabolites.

Construction of pIB061

S. erythraea NRRL2338 was transformed with pCJR30 (Rowe, C. J., et al. (1998) Gene 216:215-223) using a routine protoplast transformation technique as described by Hopwood et al. (1985). A stable integrant of *S. erythraea* [pCJR30] was identified and the production of 10mg/L of the triketide lactone (delta lactone of (2S,3R,4R,5R)-2,4-dimethyl-3,5-dihydroxy-heptanoic acid) in addition to erythromycins was confirmed by MS analysis.

Total DNA of *S. erythraea* [pCJR30] was purified and

approximately 200 ng was digested with *Eco*RI endonuclease. The digestion mixture was precipitated with isopropanol and the resulting DNA was treated with T4 DNA-ligase for 16 hours at 16°C. The ligation mixture was used to transform *E.coli* DH10B cells. The transformants were screened for the presence of the plasmid. A clone containing a 44.7kb plasmid was identified and confirmed by restriction analysis to contain three complete genes: *ery*AI, *ery*AII and *ery*AIII. The plasmid was named pIB061.

Transformation of *S. cinnamonensis*

Protoplasts of *S. cinnamonensis* were prepared by a modified procedure of Hopwood *et al.* (1985). Plasmid pIB061 was transformed into the protoplasts of *S. cinnamonensis* and stable thiostrepton resistant colonies were isolated. Individual colonies were checked for their plasmid content and the presence of plasmid pIB061 was confirmed by its restriction pattern. *S. cinnamonensis* (pIB061) was inoculated into 250 ml of M-C3 minimal production medium containing 10 μg/ml of thiostrepton and allowed to grow for 72 hours at 30 °C. After this time the mycelia were removed by filtering. The broth was extracted with two volumes of ethyl acetate and the combined ethyl acetate extracts were washed with an equal volume of saturated sodium chloride, dried over anhydrous sodium sulphate, and the ethyl acetate was removed under reduced

pressure to give about 200 mg of crude product. The product was analysed by LCQ and mass was confirmed to that of erythronolide B.

This example demonstrates the importance of *S. cinnamonensis* for production of high levels of foreign polyketide antibiotics. Introduction of the complete erythromycin gene cluster or other gene clusters into this system are likely to produce high levels of the corresponding metabolites.

Example 5

Construction of plasmid pCJW58 containing the monensin activator gene under the ermE* promoter

The ermE* promoter derived from the *ermE* resistance methyltransferase gene of *S. erythraea* (Bibb *et al.* Gene (1985) 38:215-226) was amplified by PCR as a *Spe*I-*Xba*I fragment using the following oligonucleotides 5'-CCACTAGTATGCATGCGAGTGTCCGTTCGAGT-3' and 5'-TTGTATACACCTAGGATGGTTGGCCGTGC-3' with pRH3 (Dhillon *et al.* Molecular Microbiology (1989) 3:1405-1414 as a template and cloned into *Sma*I-digested, phosphatase-treated pUC18, to produce plasmid pIB135. The integrative plasmid pSET152 (Bierman, M. *et al.* (1992) Gene 116:43-49)) was digested with *Xba*I and the backbone was dephosphorylated and ligated to the *Spe*I-*Xba*I fragment of pIB135 containing the *ermE** promoter. The ligation mixture was used to

transform *E. coli* DH10B and the orientation of the insert in the plasmids from individual clones was checked by using restriction analysis. A plasmid with the *ermE** promoter oriented so that the *Nde*I and *Xba*I sites are adjacent to the apramycin resistance gene was selected and named pIB139.

The *monR* gene from the monensin biosynthetic gene cluster was amplified and *Nde*I and *Xba*I restriction sites introduced at 5′ and 3′ ends respectively, by PCR using as primers the following oligonucleotides:
5′-AGA TAC CAT ATG CTG GGC CCG CTC CGC AT -3′
and 5′-AAT GCT CTA GAC TGT CAG CGA CCG GAC AGG GCC AA-3′
and cosmid MO.CN11 as template. The PCR product was ligated into *Sma*I-treated and phosphatase-treated plasmid pUC18 and the ligation mixture was used to transform *E. coli* DH10B cells. Transformant colonies were analysed for the presence of plasmid and the identity of the plasmid inserts was verified by sequencing. A plasmid whose insert contained the *monR* gene flanked by *Nde*I and *Xba*I restriction sites was selected and designated pCJW57.

Plasmid pCJW57 was digested with *Nde*I and *Xba*I and the fragment containing the monR gene was ligated together with the backbone of plasmid pIB139 which had been digested with the same two restriction enzymes, and purified by gel elution. The ligation mixture was used to

transform *E. coli* strain DH10B cells. Transformant colonies were analysed for the presence of plasmid and the identity of the plasmid inserts was verified by restriction analysis. One such recombinant was selected and named plasmid pCJW58.

Plasmid pCJW58 was used to transform the methylation-deficient *E. coli* strain ET 12567 (MacNeil D. J. et al. (1992) Gene 111:61-68) and the recovered, unmethylated plasmid was then used to transform the same *E. coli* strain ET12567 housing the plasmid pUB307, a derivative of RP4 which is *mob*⁻ and which contains a gene for kanamycin resistance (Piffaretti, J. C. et al. (1988) Mol. Gen. Genet. 212:215-218). Recombinants were plated on 2 x TY agar medium containing apramycin and kanamycin at final concentrations of 50 micrograms per ml and 50 micrograms per ml respectively. The plasmid content of recombinants was checked isolation of plasmid DNA and checking of the identity of these plasmids by restriction analysis. One such clone which contained both pUB307 and plasmid pCJW58 was selected and used for further experiments.

Construction of *Streptomyces cinnamonensis* (pCJW58) and production of monensins

A single colony of *E. coli* ET12567 housing both pUB307 and pCJW58 was toothpicked into 3 ml of TY liquid medium, containing apramycin and kanamycin at 25 and 25

micrograms respectively, and grown overnight at 37°C. This culture was used to inoculate 25 ml of TY medium, supplemented with the same antibiotics at the same concentrations, and growth was continued until the absorbance at 600 nm (1 cm pathlength) was between 0.3-0.6. The cells were centrifuged (room temperature, 7 minutes, 2000 x *g*), resuspended in TY liquid medium (10 ml) containing no added antibiotics, re-centrifuged as before, then resuspended in 2ml of TSB medium and placed on ice. Meanwhile, 0.5 ml of TSB medium was added to 100 microL containing approximately 10^8 spores of *S. cinnamonensis*. After a brief heat shock, at 50°C for 10 minutes, the suspension was briefly cooled, mixed with 0.5 ml of donor *E. coli* cells, and plated on solid A medium, which has composition as follows:

A medium

Sigma wheat starch	5g
Corn steep powder	1.25g
Yeast extract	1.5g
$CaCO_3$	1.5g
$FeSO_4$	6 mg
DIFCO agar	10g
H_2O	to 500 ml

pH adjusted to pH 7 with KOH.

And to which in addition was added 10 mM $MgCl_2$ to a final concentration of 10 mM.

The plates were allowed to dry overnight at room temperature, and were then allowed to incubate a further 18 hours at 30°C. After this time each 25 ml plate was overlaid with a solution of apramycin (final concentration 50 micrograms per ml) and nalidixic acid (final concentration 20 micrograms per ml), and the plates were allowed to incubate for four days at 30°C. At this time individual colonies were toothpicked onto solid A medium and allowed to grow. Four representative colonies from the A medium plate were grown up in liquid modified YEME medium, which has composition as follows:

Modified YEME medium

Sucrose	100g
DIFCO Yeast extract	3g
Bacto peptone	5g
Oxoid Malt extract	3g
Glucose	10g
H_2O to 1L	

pH adjusted to pH 7.2 with NaOH.

These cultures were used to provide a 2% vol/vol inoculum for 30 ml of modified YEME which was grown for 7 days, and then transferred to SM16 medium, which has

composition as follows:

SM16 medium

3-[N-Morpholino]-propane sulfonic acid (MOPS) buffer	20.9g
L-proline	10.0g
Glucose	20g
NaCl	0.5g
K_2HPO_4	2.1g
Ethylenediaminetetraacetic acid, sodium salt	0.25g
$MgSO_4.7H_2O$	0.49g
$CaCl_2.2H_2O$	0.029g
Trace elements solution (Hopwood, D. A. et al. (1985) Genetic Manipulation of *Streptomyces* - a Laboratory Manual, at p.235)	2 ml
0.5 M $CoCl_2$ solution	2 microlitres
H_2O to 1L	
pH adjusted to pH 7 with NaOH.	

After growth for a further 7 days, mycelium was collected by centrifugation at 2000 x g for 30 minutes, and the supernatant was extracted three times with 300 ml of ethyl acetate. The combined extracts were concentrated by evaporation under reduced pressure to an oil, which was

mixed with 1 ml of methanol. Samples were applied to an LCQ liquid chromatograph fitted with a mass spectrometer detector unit. The column used was a C18 reversed phase column, equilibrated with a mixture of 80% 20mM ammonium acetate/20% acetonitrile, and the column was eluted with a gradient of increasing acetonitrile, reaching 100% acetonitrile over 24 minutes. Monensins A and B emerged from the column with retention times respectively of 8.2 minutes and 9.2 minutes. The relative amounts of monensin produced by three independent clones (A-C) containing an additional copy of the *monR* gene were compared to a control fermentation of the wild type *S. cinnamonensis* strain, with the results shown in the Table below:

Table showing increased monensin production in strains bearing additional copy of *monR* gene

Strain	monensin A concentration (arbitrary units)	monensin B concentration (arbitrary units)
Control	188	861
A	430	1 800
B	450	1 300
C	249	1 300

Example 6

Construction of *S. cinnamonensis* M12AT5

A region lying immediately 5' of the DNA encoding the

acyltransferase (AT12) domain of module 12 of the monensin polyketide synthase in the monensin biosynthetic gene cluster was amplified with the following primers: 5'-GGTGGCCACGGAAACACCAACACCGGACCCGCGCC-3', and 5'-CTCTCGGAGGCCCGGCGCAACGGCCACAA-3', 3' using cosmid MO-CN11 as a template. The PCR product was ligated into *Sma*I digested and phosphatase-treated plasmid pUC18 and the ligation mixture was used to transform *E. coli* DH10B cells. Transformant colonies were analysed for the presence of plasmid and the identity of the plasmid inserts was verified by sequencing. A plasmid whose insert contained a fragment upstream of the AT12-encoding sequence from about 82.3kb to 83.2kb of the *mon* cluster was designated pMO81. Similarly a region lying immediately 3' of the DNA encoding the acyltransferase (AT12) domain of module 12 of the monensin polyketide synthase in the monensin biosynthetic gene cluster was amplified with the following primers: 5'-GGCCTAGGGCTGCCTCGGGTGGTGGATCTGCCGA-3' and 5'- TGGTCGGGCGCGGTGCGTGCGATACGT-3', using cosmid MO-CN11 as a template. The PCR product was ligated into *Sma*I-treated and dephosphorylated pUC18 and the ligation mixture was used to transform DH10B *E.coli* cells. Transformant colonies were analysed for the presence of plasmid and the identity of the plasmid inserts was verified by sequencing. A plasmid whose insert contained

a fragment downstream of the AT12-encoding sequence, from 80.5kb to 81.4kb of the *mon* cluster, was designated pMO82.

The DNA encoding AT of module 5 was amplified and *Msc*I and *Avr*II restriction enzyme recognition sites were introduced at the ends by PCR using the following primers: 5'-CCTGGCCAGGGCGGCCAGTGGGTGGGCATG-3' and 5'-GGCCTAGGGGTCGGCCGGGAACCAGCGCCGCCAGT-3' and the cosmid MO-CN33 as a template. The PCR product was ligated into *Sma*I-treated and dephosphorylated pUC18 and the ligation mixture was used to transform DH10B *E.coli* cells. Transformant colonies were analysed for the presence of plasmid and the identity of the plasmid inserts was verified by sequencing. A plasmid whose insert DNA, with sequence from about 44.2kb to 45.2kb of the *mon* cluster, encoded the AT5 domain was designated pMO83.

pMO81 was digested with *Msc*I and *Hind*III and ligated to the 0.9kb *Msc*I - *Hind*III fragment of pMO82. A clone containing both fragments was designated pMO84. Plasmid pMO84 was cleaved with *Avr*II and *Hind*III, treated with phosphatase, and ligated together with the 1.0 kb *Avr*II - *Hind*III fragment of pMO83 to produce pMO85, which contains the DNA encoding the AT5 domain flanked by DNA from either side of the DNA encoding the AT12 domain of the monensin PKS. The thiostrepton resistance gene *tsr*, derived from plasmid pIJ702 (Katz, E. et al., J. Gen. Microbiol.

1983), was cloned into the *Hin*dIII site of pMO85. The resulting plasmid pMO86 was analysed by its restriction pattern and confirmed to contain all the desired elements.

Plasmid pMO86 was used to transform *S. cinnamonensis* protoplasts as described by Hopwood, D. A. (1985). Stable thiostrepton-resistant transformants were isolated and checked for the desired integration of the pMO85 into the AT12 flanking regions by Southern blot hybridisation. One such integrant, *S. cinnamonensis* MO-08, containing pMO85 integrated upstream of the AT12, was passed through 4 cycles of sporulation on a non-selective nutrient medium. Spores obtained after the fourth cycle were replica-plated onto media with and without thiostrepton. DNA of clones that had lost thiostrepton resistance was analysed by Southern blot hybridisation. Clones in which the DNA encoding the AT12 domain had been replace by the DNA encoding the AT5 domain was designated *S. cinnamonensis* M12-AT5. At this time individual colonies were toothpicked onto solid A medium and allowed to grow. Four representative colonies from the A medium plate were grown up in liquid modified YEME medium, which has composition as follows:

Modified YEME medium

Sucrose	100g
DIFCO Yeast extract	3g
Bacto peptone	5g
Oxoid Malt extract	3g
Glucose	10g

H_2O to 1L

pH adjusted to pH 7.2 with NaOH.

These cultures were used to provide a 2% vol/vol inoculum for 30 ml of modified YEME which was grown for 7 days, and then transferred to SM16 medium, which has composition as follows:

SM16 medium

3-[N-Morpholino]-propane sulfonic acid (MOPS) buffer	20.9g
L-proline	10.0g
Glucose	20g
NaCl	0.5g
K_2HPO_4	2.1g
Ethylenediaminetetraacetic acid, sodium salt	0.25g
$MgSO_4.7H_2O$	0.49g
$CaCl_2.2H_2O$	0.029g
Trace elements solution (Hopwood, D. A. et al. (1985) Genetic	

Manipulation of *Streptomyces* - a Laboratory Manual, at p.235) 2 ml

0.5 M $CoCl_2$ solution 2 microlitres

H_2O to 1L

pH adjusted to pH 7 with NaOH.

After growth for a further 7 days, mycelium was collected by centrifugation at 2000 x g for 30 minutes, and the supernatant was extracted three times with 300 ml of ethyl acetate. To confirm presence of the C-2-ethyl substituents of both monensin A and B the combined extracts were concentrated by evaporation under reduced pressure to an oil, which was mixed with 1 ml of methanol. Samples were applied to an LCQ liquid chromatograph fitted with a mass spectrometer detector unit. The column used was a C18 reversed phase column, equilibrated with a mixture of 80% 20mM ammonium acetate/20% acetonitrile, and the column was eluted with a gradient of increasing acetonitrile, reaching 100% acetonitrile over 24 minutes. Mass ions 14 mass units above those expected for both monensin A and B confirmed production of the respective C-2-ethyl substituents.

Example 7. Construction of pSGK005 and its use in the production of C-13 propyl-erythromycin

Plasmid pSGK005 is a pCJR24 based plasmid containing a PKS gene comprising a loading module plus the first and

second extension modules and the chain terminating thioesterase of the PKS responsible for the production of erythromycin (DEBS). The loading module comprises the KS and ethyl-malonyl CoA specific AT from module 5 of the monensin PKS linked to the DEBS loading ACP domain. In addition, the active site cysteine of this module 5 KS has been mutated to glutamine to convert an extender di-domain to a loading di-domain. Plasmid pSGK005 was constructed as follows.

A 2769bp DNA segment of the monensin cluster of *S. cinnamonensis* extending from nucleotide 42438 to 45207 was amplified by PCR using the following oligonucleotide primers. 5'-GTGACGTCATATGTCGAGTGCTGAAGAGTCG-3' and 5'-GGGGTCGCCTAGGAACCAGCGCCGCCAGTCGA-3'

The design of these primers introduced *Nde* I and *Avr* II sites at the ends of the amplifed fragment. Monensin cosmid 05 was used as a template for the reaction. The resulting 2769bp fragment was digested with *Nde* I and *Xho* I and a 656bp fragment (Fragment A) purified by preparative gel electrophoresis.

A second PCR reaction was used with the same template to amplify DNA from nucleotide 43098 to 45207. The primers used were 5'-CGGCCTCGAGGGCCCGTCGGTCAGTGTCGACACGGCGCAGTCCTCCTCGC-3' and 5'-GGGGTCGCCTAGGAACCAGCGCCGCCAGTCGA-3'

The design of the upstream oligonucleotide primer incorporated a change of the codon specifying the KS active site cysteine (nucleotides 43135-43137, TGC) to glutamine (CAG). The resulting 2109bp DNA fragment (Fragment B) was digested with *Xho* I and *Avr* II and purified by preparative gel electrophoresis.

Plasmid pCJW80 is derived from pCJR24 and DEBS1-TE in which *Msc* I and *Avr* II sites have been introduced to flank the AT of the DEBS loading module. This plasmid was digested with *Nde* I and *Avr* II and the larger fragment (Fragment C) purified by preparative gel electrophoresis.

The three fragments (Fragments A, B, C) were ligated together using T4 DNA ligase and the ligation mixture used to transform electrocompetent *E. coli* DH10B cells. Individual clones were checked for the presence of the desired plasmid pSGK005. The identity of pSGK005 was confirmed by restriction pattern and sequence analysis.

Plasmid pSGK005 was used to transform *S. erythraea* NRRL2338 using a routine protoplast transformation technique. Thiostrepton resistant colonies were selected on R2T20 media containing g/ml thiostrepton. Further analysis confirmed that pSGK005 had integrated into the *S. erythraea* NRRL2338 chromosome by Southern blot hybridisation of their genomic DNA with DIG-labelled DNA containing the *act*II orf4 promoter. The culture *S.*

erythraea NRRL2338 (pSGK005) was inoculated into 5ml tap water medium in a 30ml flask. After three days incubation at 29°C this flask was used to inoculate 30ml of Ery-P medium in a 300ml flask. The broth was incubated at 29°C at 200rpm for 6 days. After this time the whole broth was adjusted to pH8.5 with NaOH, and then extracted twice with an equal volume of ethyl acetate. The ethyl acetate extract was evaporated to dryness at 45°C under a nitrogen stream using a Zymark Turbovap LV evaporator. The product identities were confirmed by LC/MS. A peak was observed with a m/z value of 734 $(M+H)^+$ required for erythromycin A. A second peak was observed with a m/z value of 748 $(M+H)^+$, required for 13-propyl erythromycin A.

References

1. Ajaz, A.A. and Robinson, J.A. (1983) The utilization of oxygen atoms from molecular oxygen during the biosynthesis of Monensin-A. *Journal of the Chemical Society-Chemical Communications*, **12**, 679-680.

2. Ashworth, D.M., Holmes, D.S., Robinson, J.A., Oikawa, H. and Cane, D.E. (1989) Selection of a specifically blocked mutant of *Streptomyces cinnamonensis* - isolation and synthesis of 26-deoxymonensin-A. *Journal of Antibiotics*, **42**, 1088-1099.

3. August, P.R., Tang, L., Yoon, Y.J., Ning, S., Muller, R., Yu, T.W., Taylor, M., Hoffmann, D., Kim, C.G., Zhang, X.H., Hutchinson, C.R. and Floss, H.G. (1998) Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chemistry & Biology*, **5**, 69-79.

4. Bartel, P.L., Zhu, C.B., Lampel, J.S., Dosch, D.C., Connors, N.C., Strohl, W.R., Beale, J.M. and Floss, H.G. (1990) Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in *Streptomycetes* - clarification of actinorhodin gene functions. *Journal of Bacteriology*, **172**, 4816-4826.

5. Bibb, M.J., Biro, S., Motamedi, H., Collins, J.F. and Hutchinson, C.R. (1989) Analysis of the nucleotide sequence of the *Streptomyces glaucescens Tcml* genes provides key information about the enzymology of polyketide antibiotic

biosynthesis. *EMBO Journal*, **8**, 2727-2736.

6. Bibb, M.J. and Hopwood, D.A. (1981) Genetic studies of the fertility plasmid SCP2 and its SCP2* variants in *Streptomyces coelicolor* A3(2). *Journal of General Microbiology*, **126**, 427-442.

6a. Bibb M.J., Janssen G.R. and Ward J.M. (1985) Cloning and analysis of the promoter region of the erythromycin resistance gene (ErmE) of *Streptomyces erythraeus*. *Gene*, **38**, 215-226.

6b. Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao R.N. and Schoner B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene*, **116**, 43-49.

7. Cane, D.E., Liang, T.C. and Hasler, H. (1981) Polyether biosynthesis - origin of the oxygen atoms of Monensin-A. *Journal of the American Chemical Society*, **103**, 5962-5965.

8. Cane, D.E., Liang, T.C. and Hasler, H. (1982) Polyether biosynthesis 2. Origin of the oxygen atoms of monensin A. *Journal of the American Chemical Society*, **104**, 7274-7281.

9. Cortés, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F. (1990) An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature*, **348**, 176-178.

10. Cortés, J., Wiesmann, K.E.H., Roberts, G.A., Brown, M.J.B., Staunton, J. and Leadlay, P.F. (1995) Repositioning

of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science*, **268**, 1487-1489.

11. Davis, N.K. and Chater, K.F. (1990) Spore color in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Molecular Microbiology*, **4**, 1679-1691.

12. Day, L.E. (1973) *Antimicrobial Agents and Chemotherapy*, **4**, 410-414.

13. Demetriadou, A.K., Laue, E.D., Staunton, J., Ritchie, G.A.F., Davies, A. and Davies, A.B. (1985) Biosynthesis of the polyketide polyether antibiotic ICI-139603 in *Streptomyces longisporoflavus* from O-18-labeled acetate and propionate. *Journal of the Chemical Society Chemical Communications*, **7**,408-410.

13a. Dhillon, N., Hale, R.S., Cortes, J. and Leadlay P.F. (1989) Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. *Molecular Microbiology* **3**, 1405-1414.

14. Donadio, S., McAlpine, J.B., Sheldon, P.J., Jackson, M. and Katz, L. (1993) An erythromycin analog produced by reprogramming of polyketide synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 7119-7123.

15. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J. and Katz, L. (1991) Modular organization of genes required

for complex polyketide biosynthesis. *Science*, **252**, 675-679.

16. Evans, G.A., Lewis, K. and Rothenberg, B.E. (1989) High efficiency vectors for cosmid microcloning and genomic analysis. *Gene*, **79**, 9-20.

17. Fernandez-Moreno, M.A., Caballero, J.L., Hopwood, D.A. and Malpartida, F. (1991) The Act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* transfer-RNA gene of *Streptomyces*. *Cell*, **66**, 769-780.

18. Fernandez-Moreno, M.A., Martinez, E., Boto, L., Hopwood, D.A. and Malpartida, F. (1992) Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *Journal of Biological Chemistry*, **267**, 19278-19290.

19. Gramajo, H.C., Takano, E. and Bibb, M.J. (1993) Stationary phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. *Molecular Microbiology*, **7**, 837-845.

20. Gumila, C., Ancelin, M.L., Delort, A.M., Jeminet, G. and Vial, H.J. (1997) Characterization of the potent *in vitro* and *in vivo* antimalarial activities of ionophore compounds. *Antimicrobial Agents and Chemotherapy*, **41**, 523-529.

21. Hallam, S.E., Malpartida, F. and Hopwood, D.A. (1988) Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor*. *Gene*, **74**, 305-320.

22. Holmes, D.S., Sherringham, J.A., Dyer, U.C., Russell, S.T. and Robinson, J.A. (1990) Synthesis of putative intermediates on the monensin biosynthetic pathway and incorporation experiments with the monensin-producing organism. *Helvetica Chimica Acta*, **73**, 239-259.

23. Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985) *Genetic manipulation of Streptomyces, a laboratory manual*. John Innes Institution, Norwich, UK.

24. Hutchinson, C.R. and Fujii, I. (1995) Polyketide synthase gene manipulation - a structure function approach in engineering novel antibiotics. *Annual Review of Microbiology*, **49**, 201-238.

25. Hutchinson, C.R., Sherman, M.M., Vederas, J.C. and Nakashima, T.T. (1981) Biosynthesis of macrolides .5. Regiochemistry of the labeling of lasalocid a by C-13,O-18-labeled precursors. *Journal of the American Chemical Society*, **103**, 5953-5956.

26. Kakavas, S.J., Katz, L. and Stassi, D. (1997) Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*. *Journal of Bacteriology*, **179**, 7515-7522.

27. Kao, C.M., Luo, G.L., Katz, L., Cane, D.E. and Khosla, C. (1995) Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase. *Journal of the American Chemical Society*, **117**, 9105-9106.

28. Katz, E., Thompson, C.J. and Hopwood, D.A. (1983) Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. *Journal of General Microbiology*, **129**, 2703-2714.

29. MacNeil, D.J., Occi, J.L., Gewain, K.M., Macneil, T., Gibbons, P.H., Ruby, C.L. and Danis, S.J. (1992) Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. *Gene*, **115**, 119-125.

29a. MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons,P.H. and MacNeil, T. (1992) Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61-68.

30. Marsden, A.F.A., Wilkinson, B., Cortés, J., Dunster, N.J., Staunton, J. and Leadlay, P.F. (1998) Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science*, **279**, 199-202.

31. Meurer, G., Gerlitz, M., Wendt Pienkowski, E., Vining, L.C., Rohr, J. and Hutchinson, C.R. (1997) Iterative type II polyketide synthases, cyclases and ketoreductases exhibit context-dependent behavior in the biosynthesis of linear and angular decapolyketides. *Chemistry & Biology*, **4**, 433-443.

32. Motamedi, H. and Shafiee, A. (1998) The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *European Journal of Biochemistry*, **256**, 528-534.

33. Narva, K.E. and Feitelson, J.S. (1990) Nucleotide sequence and transcriptional analysis of the RedD locus of

Streptomyces coelicolor A3(2). *Journal of Bacteriology*, **172**, 326-333.

33a. Piffaretti J.C., Arini A. and Frey J. (1988) pUB307 mobilizes resistance plasmids from Escherichia coli into Neisseria gonorrhoeae. *Mol Gen Genet*.**212**, :215-218.

34. Pospisil, S., Sedmera, P., Vokoun, J., Vanek, Z. and Budesinsky, M. (1987) 3-O-Demethylmonensin-A and 3-O-demethylmonensin-B produced by *Streptomyces cinnamonensis*. *Journal of Antibiotics*, **40**, 555-557.

35. Revill, W.P., Bibb, M.J. and Hopwood, D.A. (1995) Purification of a malonyltransferase from *Streptomyces coelicolor* A3(2) and analysis of its genetic determinant. *Journal of Bacteriology*, **177**, 3946-3952.

36. Rowe, C.J., Cortés, J., Gaisser, S., Staunton, J. and Leadley, P.F. (1998) Construction of new vectors for high-level expression in actinomycetes. *Gene*, **216**, 215-223.

37. Sanger, F. (1981) Determination of nucleotide sequences in DNA. *Science*, **214**, 1205-1210.

38. Schwecke, T., Aparicio, J.F., Molnár, I., Konig, A., Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J.B., Böhm, G.A., Staunton, J. and Leadlay, P.F. (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 7839-7843.

39. Shen, B., Summers, R.G., Wendtpienkowski, E. and Hutchinson, C.R. (1995) The *Streptomyces glaucescens* TcmKl polyketide synthase and TcmN polyketide cyclase genes govern the size and shape of aromatic polyketides. *Journal of the American Chemical Society*, **117**, 6811-6821.

40. Sherman, D.H., Malpartida, F., Bibb, M.J., Kieser, H.M. and Hopwood, D.A. (1989) Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. *EMBO Journal*, **8**, 2717-2725.

41. Sood, G.R., Robinson, J.A. and Ajaz, A.A. (1984) Biosynthesis of the polyether antibiotic Monensin-A - incorporation of [2-2-2H-2]-propionate, (R)-[2-2H-1]-propionate and (S)-[2-2H-1]- propionate. *Journal of the Chemical Society-Chemical Communications*, **21**, 1421-1423.

42. Spavold, Z., Robinson, J.A. and Turner, D.L. (1986) Biosynthesis of the polyether antibiotic narasin origins of the oxygen atoms and the mechanisms of ring formation. *Tetrahedron Letters*, **27**, 3299-3302.

43. Staden, R. (1996) The Staden sequence analysis package. *Molecular Biotechnology*, **5**, 233-241.

44. Stutzman-Engwall, K.J., Otten, S.L. and Hutchinson, C.R. (1992) Regulation of secondary metabolism in *Streptomyces* Spp and overproduction of daunorubicin in *Streptomyces peucetius*. *Journal of Bacteriology*, **174**, 144-154.

45. Swan, D.G., Rodriguez, A.M., Vilches, C., Mendez, C. and Salas, J.A. (1994) Characterization of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which

has an unusual coding sequence. *Molecular & General Genetics*, **242**, 358-362.

46. Takano, E., Gramajo, H.C., Strauch, E., Andres, N., White, J. and Bibb, M.J. (1992) Transcriptional regulation of the *redD* transcriptional activator gene accounts for growth phase-dependent production of the antibiotic undecylprodigiosin in *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, **6**, 2797-2804.

47. Townsend, C.A. and Basak, A. (1991) Experiments and speculations on the role of oxidative cyclization chemistry in natural product biosynthesis. *Tetrahedron*, **47**, 2591-2602.

48. Walba, D.M. and Edwards, P.D. (1980) *Tetrahedron Letters*, **21**, 3531-3534.

49. Westley, J.W. (1974) *Journal of Antibiotics*, **27**, 597-604.

50. Westley, J.W. (1981) *Antibiotics IV. Biosynthesis*. Springer-Verlag, New York.

51. Wietzorrek, A. and Bibb, M. (1997) A novel family of proteins that regulates antibiotic production in *Streptomycetes* appears to contain an OmpR-like DNA-binding fold. *Molecular Microbiology*, **25**, 1181-1184.

52. Xue, Y.Q., Zhao, L.S., Liu, H.W. and Sherman, D.H. (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 12111-12116.

TABLE I

gene	function	start	end
gdhA	glutamate dehydrogenase (partial)	1038	0
dapA	dihydrodipicolinate synthase	2140	1220
orf3	putative transcriptional activator	2211	3152
orf4	hypothetical protein	3264	3680
orf5	hypothetical protein	4307	3684
orf6	hypothetical protein	4570	4758
orf7	hypothetical protein	5058	5612
acpX	acyl carrier protein	6010	5693
ksX	ketoacyl synthase	8531	6045
monCI	probable epoxihydrolase/cyclase	9542	8643
monE	methyltransferase	10426	9596
monT	monensin resistance gene (ABC-	10656	12191
monRI	probable repressor	12205	12780
monAI	thioesterase	13829	13023
monAI	polyketide synthase loading &	14121	23198
	KS-L	14172	15486
	AT-L malonate specific	15777	16880
	ACP-L	17019	17276
	KS1	17358	18626
	AT1 methylmalonate specific	18960	19976
	DH1 (potential)	20019	20519
	KR1 (inactive)	21636	22241
	ACP1	22536	22793
monAI	polyketide synthase module 2	23205	29921
	KS2	23307	24569
	AT2 methylmalonate specific	24891	25913
	DH2	25953	26369
	ER2	27600	28463
	KR2	28485	29042
	ACP2	29313	29570
monAI	polyketide synthase modules 3 & 4	29974	42372
	KS3	30076	31347
	AT3 malonate specific	31798	32838
	DH3	32884	33465
	KR3	34692	35181
	ACP3	35553	35811
	KS4	35899	37170
	AT4 methylmalonate specific	37489	38511
	DH4	38557	38982
	ER4	40123	40986
	KR4	41005	41562
	ACP4	41848	42105
monAI	polyketide synthase modules 5 & 6	42448	54564
	KS5	42628	43890
	AT5 ethylmalonate specific	44221	45243
	DH5	45289	45744
	KR5	46785	47337
	ACP5	47593	47850

	KS6	47947	49218
	AT6 malonate specific	49579	50601
	DH6	50644	51075
	ER6	52222	53102
	KR6	53101	53661
	ACP6	54052	54306
monA	polyketide synthase modules 7 & 8	54614	66934
	KS7	54716	55978
	AT7 methylmalonate specific	56300	57319
	DH7	57358	57802
	KR7	59048	59608
	ACP7	59867	60124
	KS8	60185	61453
	AT8 malonate specific	61808	62839
	DH8	62882	63316
	ER8	64577	65437
	KR8	65456	66016
	ACP8	66404	66661
monA	polyketide synthase module 9	66952	72054
	KS9	67075	68340
	AT9 malonate specific	68698	69729
	KR9 (potential)	70735	71262
	ACP9	71536	71783
monH	probable regulator	72051	74993
monCI	FAD containing epoxidase	76541	75051
monBI	double bond isomerase	76960	76538
monBI	double bond isomerase	77450	77016
monA	polyketide synthase modules 11 &	88708	77447
	KS11	88612	87344
	AT11 methylmalonate specific	87022	85993
	KR11	85111	84562
	ACP11	84292	84035
	KS12	83962	82694
	AT12 methylmalonate specific	82354	81335
	DH12 (potential) delta	81286	80855
	ER12 (potential)	79618	78914
	KR12	78895	78337
	ACP12	78070	77812
monA	polyketide synthase module 10	93741	88816
	KS10	93636	92368
	AT10 methylmalonate specific	92040	91021
	KR10	90132	89584
	ACP10	89322	89068
monD	P450 oxygenase	94081	95273
monRI	probable activator	96141	95338
monA	thioesterase	96941	96138
orf29	cell wall biosynthesis capK	97580	98953
lipB	lipase B	99983	98991
orf31	ion pump	101433	100507
orf32	membrane structural protein	102581	101490
amtA	glycine amidinotransferase	102924	103450

TABLE II

GdhA, glutamate dehydrogenase (partial coding sequence) Length: 346 amino acids

```
  1  LTTRPDTKTA LSQKTALSQL LTEIEHRNPA QPEFHQAARE VLETLAPVIA
 51  ARPEYAEAGL IERLCEPERQ IVFRVPWQDD HGRVRVNRGF RVEFNSALGP
101  YKGGLRFHPS VNLGVIKFLG FEQIFKNALT GLGIGGGKGG SDFDPRGRSD
151  AEVMRFCQSF MTELYRHIGE HTDVPAGDIG VGGREIGYLF GQYRRITNRW
201  EAGVLTGKGR NWGGSLIRPE ATGYGNVLFA AAMLRERGET LEGRTAVVSG
251  SGNVAIYTIQ KLAALGANAV TCSDSSGYVV DEKGIDLDLL KQVKEVERAR
301  VDTYAQRRGA SARFVPGRRV WEVPADIALP SATQNELDAD DATALI
```

DapA, dihydrodopicolinate synthase Length: 307 amino acids

```
  1  MTLASSLEPT TEPLFNGLYV PLVTPFTDDL RLAPEALARL ADEALSAGAS
 51  GLVALGTTAE AATLTAEERE TVIRVCSAAC RAHGAPLIVG VGTNDTATAI
101  TALRELAARG DVAAALVPAP PYIRPGEAGT LAHFAALAEH GGLPLVVYDI
151  PYRTGQTLGA GTITALGRLP EVVGIKHATG SIDPTTMELL DSPLPGFAVL
201  GGDDIVLSPL VAAGAHGGIV ASANLRTADY AEMIALWRRG SAAPARALGA
251  DLARLSAALF TEPNPTVIKG VLHAQNRIPS PAVRMPLLAA SADSVRRAAP
301  LAASRK*
```

ORF3, putative transcriptional activator protein Length: 314 amino acids

```
  1  MLDVRRLHLL RELDRRGTIA AVAEALTFTA SAVSQQLGVL EREAGVPLLE
 51  RSGRRVVLTP AGRSLVAHAD AVLNRLEQAV AELAGARDGI GGPLRIGTFP
101  SGGHTIVPGA LAELASRHPA LEPMVREIDS ARVSDGLRAG ELDVALVHDY
151  DFVPATPDTT VDEVPLLEEP MYLVTHAADT ATDSGSGSTL AALLGPCAEV
201  PWITARDGTT GHAMAVRACQ AAGFQPRIRH QVNDFRTVLA LVAAGQGAGF
251  VPRMAAEPSP AGVVLTKLPL FRRSKVAFRA GGGAHPAIAA FVAAATTAVE
```

301 RMAGSRGPAG GSE*

ORF4, hypothetical protein Length: 139 amino acids

1 MADDAYLFLL PDRHPRLGAA LAAVGALECT ETPAVHAWLQ AHEASVSSEQ
51 VRILPADAET LIPKDAERLP VPLSEEEALK VEQECAPQTV TDMESELLAF
101 RETTQDWQAL VHRALTAGIP AQRIARLTGL DPEEIGRL*

ORF5, hypothetical protein Length: 208 amino acids

1 LAVAACAAVV LPIDAVVRIS AADVGVLVFF AYLLPYLAIT MTVFVSVAPE
51 QVRSWARREA RGTFLQRYVL GTAPGPGGSL FIAAAALVVA VLWLPGHLST
101 TFSALPRTLV ALALVVAAWI CVVVAFAVTF QADNLVENER ALEFPGERSP
151 AWADYVYFAL AAMTTFGTTD VDVTSRDMRR TVAANTVIAF VFNTVTVAIL
201 VSALGGR*

ORF6, hypothetical protein Length: 63 amino acids

1 MTVMDKLKQM LKGHEDKAGQ GIDKAGDFVD GKTQGKYSGQ VDTAQDKLRD
51 QFGSDQQEPP QR*

ORF7, hypothetical protein Length: 185 amino acids

1 MGTAQSQEQA AAPGACAAFV RFVLCGGGVG LASSFAVVAL ASWVPWALAN
51 ALVAVVSTVV ATELHARFTF GAGGRATWRQ HAQSAGSAAA AYAVTCVAMF
101 VLQQLVAAPG AVLEQVVYLS ASALAGVARF VVLRLVVFAR NRSLPAAAAV
151 RTARPVRRVP APVPATVAHA ASRPAGPAAL CPAA*

AcpX, acyl carrier protein (ACP) Length: 106 amino acids

1 MTSTDHTSGQ DATELEKQLA AATPEEREKL LTDTIRTQAG TLLNTTLSDD
51 SNFLENGLNS LTALELTKTL MTLTGMEIAM VAIVENPTPA QLAHHLGQEL
101 AHTTA*

KsX, ketoacyl-ACP synthase Length: 829 amino acids

```
  1  VANEEKLVEY LKWTTAELHQ AQQQLRELKA AQHEPIAVVS MACRLPGKTR
 51  TPDDLWDLVS EGRDAVTGFP DDRAWELPEE RPYAELGGFL DDAAGFDAGF
101  FDISDTEAVA TEPLQRLMLH LAWETVERGH IAPHTLRSTL TGVYVGATGH
151  DYATRLETAP DELLPYLGGG TSGSLVSGRI AYALGLEGPA ISVDTACSSS
201  LVALHLACQA LRRGECGLAL AGGGTVMSTP HTFHAFAHQK SLAQDGRCKP
251  FAAAADGMGL GEGVGLVLLE RLGDARKNGH PVLAVIRGSA VNQDGAGYGL
301  AAPNGPSQQH VIRAALADAG LTPDQIDAVE AHGTGTPIGD AIEVQALLAT
351  YGADRSPDRP LWLGSVKSNT GHTQGAAGAA ALIKMVQAFR HGTLPPTLHV
401  DRPTPLAAWK KGAVRLLTEA VDWPRREEPR RVGISAFATS GTNAHLILEE
451  PPVDEAPVPD AARDQTSPVA PELPVAWSLS ARTPEALRAQ AKALVTHLAA
501  TDPAPSPAEV AYSLAATRSP LEHRAVLTGT DHTELLAAAR ALAAGEDHPD
551  LVRSTPGAGP KKIAWHFDGR PADGVTTGAA PGAKPGATFG ATFGAAFGGA
601  EFHSAFPLFA SAFDEARALL DTHLPTPLPT PHSELARFAV HTALARLLLE
651  TGVRPHTLTG DGVGHIAAAY AAGILTLDDA CRLAAAHAAA AQAAEGEQPA
701  PPDAYEPVLK QLTFQRATLT LTSTAPADTP IASADYWHHH LTSPAPTAPP
751  TPETHTLLHL GALSPEGTQT SAVSALLTAL ARLHTTGGTV DWTPLVRRTP
801  HPRTIDLPTY SFQATRYWLH DHTAHAAV*
```

MonCII, probable epoxyhydrolase/cyclase Length: 300 amino acids

```
  1  VKNLRIPVSQ TVSLNVRYRP ADGPGAPGRP FLLLHGMLSN ARMWDEVAAR
 51  LAAAGHPAYA VDHRGHGESD TPPDGYDNAT VVTDLVAAVT ALDLSGALVA
101  GHSWGAHLAL RLAAEHPDLV AGLALIDGGW YEFDGPVMRA FWERTADVVR
151  RAQQGTTSAA DMRAYLRATH PDWSPTSIEA RLADYRVGPD GLLIPRLTST
201  QVMSIVAGLQ REAPADWYPK VTVPVRLLPL IPAIPQLSDQ VRAWVAAAEA
```

```
251  ALEQVSVRWY PGSDHDLHAG APDEIAADLL LLARSCEAMP GGKAGVRPA*
```

MonE, S-adeonosylmethionine-dependent methyltransferase Length: 277 amino acids

```
  1  VNKTVAPEPS DIGHYYDHKV FDLMTQLGDG NLHYGYWFDG GEQQATFDEA
 51  MVQMTDEMIR RLDPAPGDRV LDIGCGNGTP AMQLARARDV EVVGISVSAR
101  QVERGNRRAR EAGLADRVRF EQVDAMNLPF DDGSFDHCWA LESMLHMPDK
151  QQVLTEAHRV VKPGARMPIA DMVYLNPDPS RPRTATVSDT TIYAALTDIG
201  DYPDIFRAAG WTVLELTDIT RETAKTYDGY VEWIRAHRDE YVDIIGVEGY
251  ELFLHNQAAL GKMPELGYIF ATAQRP*
```

MonT, putative monensin resistance gene (ABC-transporter) Length: 512 amino acids

```
  1  MSADLGARRW WAVGALVLAS MVVGFDVTIL SLALPAMADD LGANNVELQW
 51  FVTSYTLVFA AGMIPAGMLG DRFGRKKVLL TALVIFGIAS LACAYATSSG
101  TFIGARAVLG LGAALIMPTT LSLLPVMFSD EERPKAIGAV AGAAMLAYPL
151  GPILGGYLLN HFWWGSVFLI NVPVVILAFL AVSAWLPESK AKEAKPFDIG
201  GLVFSSVGLA ALTYGVIQGG EKGWTDVTTL VPCIGGLLAL VLFVMWEKRV
251  ADPLVDLSLF RSARFTSGTM LGTVINFTMF GVLFTMPQYY QAVLGTDAMG
301  SGFRLLPMVG GLLVGVTVAN KVAKALGPKT AVGIGFALLA AALFYGATTD
351  VSSGTGLAAA WTAAYGLGLG IALPTAMDAA LGALSEDSAG VGSGVNQSIR
401  TLGGSFGAAI LGSILNSGYR GKLDLDGVPE QAHGAVKDSV FGGLAVARAI
451  KSNGLADSVR SAYVHALDVV LVVSGGLGLL GVVLAVVWLP RHVGQSTAKT
501  AESEHEAADA V*
```

MonRII, probable repressor protein Length: 192 amino acids

```
  1  VPGLRERKKA RTKAAIQREA VRLFREQGYT ATTIEQIAEA AEVAPSTVFR
 51  YFATKQDLVF SHDYDLPFAM MVQAQSPDLT PIQAERQAIR SMLQDISEQE
```

101 LALQRERFVL ILSEPELWGA SLGNIGQTMQ IMSEQVAKRA GRDPRDPAVR

151 AYTGAVFGVM LQVSMDWAND PDMDFATTLD EALHYLEDLR P*

MonAIX, thioesterase Length: 269 amino acids

1 MDRGTAARAP QIGDEFGAAT GNGVWLRRYH AAAEAPVRLV CFPFAGGSAS

51 YYFGLSGLLA PGVEVLAVQY PGRQDRHAEP CLASVAELAD GVVPHLPCDG

101 KPFALFGHSL GAIVAFEVAR RLRGPAGPGL PVHLFVSGGL ARPYRPAGRS

151 GAFGDADILA HLRAMGGTDE RFFRSPELQE LVLPALRADY RAVATYEAPG

201 PGRLDCPITA LIGDADERTS PEQAATWRER TGAAFDLRVL PGGHFYLDGC

251 QEQVAAVVTE ALTAGPGV*

MonAI, polyketide synthase multi-enzyme MONS1, housing loading module and extension module 1 Length: 3026 amino acids

1 MAASASASPS GPSAGPDPIA VVGMACRLPG APDPDAFWRL LSEGRSAVST

51 APPERRRADS GLHGPGGYLD RIDGFDADFF HISPREAVAM DPQQRLLLEL

101 SWEALEDAGI RPPTLARSRT GVFVGAFWDD YTDVLNLRAP GAVTRHTMTG

151 VHRSILANRI SYAYHLAGPS LTVDTAQSSS LVAVHLACES IRSGDSDIAF

201 AGGVNLICSP RTTELAAARF GGLSAAGRCH TFDARADGFV RGEGGGLVVL

251 KPLAAARRDG DTVYCVIRGS AVNSDGTTDG ITLPSGQAQQ DVVRLACRRA

301 RITPDQVQYV ELHGTGTPVG DPIEAAALGA ALGQDAARAV PLAVGSAKTN

351 VGHLEAAAGI VGLLKTALSI HHRRLAPSLN FTTPNPAIPL ADLGLTVQQD

401 LADWPRPEQP LIAGVSSFGM GGTNGHVVVA AAPDSVAVPE PVGVPERVEV

451 PEPVVVSEPV VVPTPWPVSA HSASALRAQA GRLRTHLAAH RPTPDAARVG

501 HALATTRAPL AHRAVLLGGD TAELLGSLDA LAEGAETASI VRGEAYTEGR

551 TAFLFSGQGA QRLGMGRELY AVFPVFADAL DEAFAALDVH LDRPLREIVL

601 GETDSGGNVS GENVIGEGAD HQALLDQTAY TQPALFAIET SLYRLAASFG

651 LKPDYVLGHS VGEIAAAHVA GVLSLPDASA LVATRGRLMQ AVRAPGAMAA

701 WQATADEAAE QLAGHERHVT VAAVNGPDSV VVSGDRATVD ELTAAWRGRG

751 RKAHHLKVSH AFHSPHMDPI LDELRAVAAG LTFHEPVIPV VSNVTGELVT

801 ATATGSGAGQ ADPEYWARHA REPVRFLSGV RGLCERGVTT FVELGPDAPL

851 SAMARDCFPA PADRSRPRPA AIATCRRGRD EVATFLRSLA QAYVRGADVD

901 FTRAYGATAT RRFPLPTYPF QRERHWPAAA GVGQQPETPE LPESSESSEQ

951 AGHEREEGAR AWGGPEGRLA GLSVNDQERV LLGLVTKHVA VVLGDASGTV

1001 QAARTFKQLG FDSMAAAELS ERLGTETGLP LPATLTFDYP TPLAVAAHLR

1051 AELTGTPAPA GSAPATGALG AGDLGTDEDP VAIVAMSCRY PGGAGTPEDL

1101 WRLVADGADA IGDFPTDRGW DLARLFHPDP DRSGTSCTRQ GGFLYDAADF

1151 DAEFFDISPR EALAVDPQQR LLLECAWEAF ERAGLDPRAL KGSPTGVFVG

1201 MTGQDYGPRL HEPSQATDGY LLTGSTPSVA SGRLSFSFGL EGPALTVDTA

1251 CSSSLVTLHL AAQALRRGEC DLALAGGATV LATPGMFTEF SRQRGLAPDG

1301 RCKPFAAGAD GTGWAEGVGL VLLERLSEAR RKGHAVLAVI RGSAINQDGA

1351 SNGLTAPNGP SQQRVIRAAL AAARLTADEV DVVEAHGTGT TLGDPIEAQA

1401 LLATYGQGRS AERPLWLGSV KSNIGHTQAA AGVAGVIKMV MAMRHDLLPA

1451 TLHVDEPSGH VDWSTGAVRL LTEPVVWPRG ERPRRAAVSS FGISGTNAHL

1501 VLEEAGQDEY VAGAADDAGP VDGAVLPWVV SGRTGAALRE QARRLRELVT

1551 GGSADVSVSG VGRSLVTTRA VFEHRAVVVG RDRDTLIGGL EALAAGDASP

1601 DVVCGVAGDV GPGPVLVFPG QGSQWVGMGA QLLGESAVFA ARIDACEQAL

1651 SPYVDWSLTE VLRGDGRELS RVDVVQPVLW AVMVSLAAVW ADHGVTPAAV

1701 VGHSQGEIAA VVVAGALTLE DGAKIVALRS RALRQLSGGG AMASLGVGQE

1751 QAAELVEGHP GVGIAAVNGP SSTVISGPPE QVAAVVADAE ARELRGRVID

1801 VDYASHSPQV DAITDELTHT LSGVRPTTAP VAFYSAVTGT RIDTAGLDTD

1851 YWVTNLRRPV RFADAVTALL ADGHRVFIEA SSHPVLTLGL QETFEEAGVD
1901 AVTVPTLRRE DGGRARLARS LAQAFGAGCA VRWENWFPAT GTSTVELPTY
1951 AFQRRRYWLE APTGTQDAAG LGLAAAGHPL LGAATEIADG DIRLLTGRIS
2001 RHSHPWLAQH TLFGAAVVPA SVLAEWALRA ADEAGCPRVD DLTLRTPLVL
2051 PETAGVQVQI VVGPADARDG HRDFHVYARP DGKDASEGEG IAEGEGASEG
2101 EGASGGTDAP WTCHADGRLV AEPTGTASED SPDTVWPPPG AEPVDLGDFY
2151 ERAAATGVGY GPVFTGLRAL WRRDGELFAE AVLPQEAPET AGFGMHPALL
2201 DAALHPALLG ERPAEEDKVW LPFTLTGVTL WATGATSVRV RLTPLDDDPD
2251 ASADGRAWRV GVSDPTGAEV LTCEALVAVA AGRRELRAAG ERVSDLYAVE
2301 WVPVPGPGPV GEGADFSGWA GLGECGERWE CVGRVERWYE DLDALGAAVE
2351 GGASVPSVVL ATAAAAPGGA GDGAADALSA VRWTGALLDQ WLADARFADA
2401 RLVVITSGAV ATGDDFLPDP AAAAVRGLVE QAQVRHPGRI LLVDTEAGAG
2451 LGVGAGVDDA LLEQAVAMAL GADEPQLALR AGRVLAPRLT APQDAAVTEA
2501 ARPLDPDGTV LITGPAGAPV ADLAEHLVRT GQCRHLLLLP GDGELEEMAE
2551 ELRGLGATVD LSTADPADPT ALAEVVAAVE GDHPLTGVIH ATGVVDAFDP
2601 GDSASDLMID SASDSFAEAW SSRAGVTAAL HTATAHLPLD LFAVLSPAGA
2651 DLGIARSAAA AGADAFSAAL ALRRHTTVTT DTTAPPRTTA PPRTTASPRT
2701 TALSSSRTTG VALAYGPPTA PRPGIKGTAP GRIPVLLDAA RAHGGGSPLL
2751 GARLAARALA AESAAEGVAG LPAPLRALAV AAAAAGAPTR RTAADRKPPA
2801 DWPARLAPLS APEQLRLLID AVRTHAAAVL GRTDPEALRG DATFKQLGLD
2851 SLTAVELRNR LVEDTGLRLP TALVFRYPTP AAIAAHLRER LTSPSETTAT
2901 QRSGGQTPAA GQASSALAPG GSAAGPPAAD TVLSDLTRME NTLSVLAAQL
2951 PHTETGEITT RLEALLTRWK TTNATANDSG DGNGGDDDAA ERLKAASADQ
3001 IFDFIDNELG VGHGTSRVTP TPKAG*

MonAII, polyketide synthase multi-enzyme MONS2, housing extension module 2 Length: 2239 amino acids

```
   1  MASEEQLVEY LRRVTTELHD TRRRLVQEED RRQEPVALVG MACRFPGGVA
  51  SPEDLWDLVA AGKDAIEDFP TDRGWDLEAL YDPDPAAYGT SYVRHGGFVD
 101  DAGSFDADFF GISPREALAM DPQQRLMLET SWELFERAGI EPVSLKGSRT
 151  GVYAGVSSED YMSQLPRIPE GFEGHATTGS LTSVISGRVA YNYGLEGPAV
 201  TVDTACSASL VAIHLASQAL RQRECDLALA GGVLVLSSPL MFTEFCRQRG
 251  LAPDGRCKPF AAAADGTGFS EGIGLLLLER LSDARRNGHK VLAVIRGSAV
 301  NQDGASNGLT APNDAAQEQV IRAALDNARL TPSEVDAVEA HGTGTKLGDP
 351  IEAGALLATY GQHRARPLLL GSLKSNIGHT HATAGVAGVI KTVMAIRNGL
 401  LPATLHVEEL SPHVDWDAGA VEVVTEPTPW PETGHPRRAG VSAFGISGTN
 451  AHLILEEAPP EEDVPAPVVV ESGGVVPWVV SGRTPEALRE QARRLGEFVA
 501  GDTDALPNEV GWSLATTRSV FEHRAVVVGR DRDALTAGLG ALAAGEASAG
 551  VVAGVAGDVG PGPVLVFPGQ GAQWVGMGAQ LLDESAVFAA RIAECERALS
 601  AHVDWSLSAV LRGDGSELSR VEVVQPVLWA VMVSLAAVWA DYGVTPAAVI
 651  GHSQGEMAAA CVAGALSLED AARIVAVRSD ALRQLQGHGD MASLSTGAEQ
 701  AAELIGDRPG VVVAAVNGPS STVISGPPEH VAAVVADAEA RGLRARVIDV
 751  GYASHGPQID QLHDLLTERL ADIRPTNTDV AFYSTVTAER LTDTTALDTD
 801  YWVTNLRQPV RFADTIEALL ADGYRLFIEA SAHPVLGLGM EETIEQADMP
 851  ATVVPTLRRD HGDTTQLTRA AAHAFTAGAD VDWRRWFPAD PAPRTIDLPT
 901  YAFQRRRYWL ADTVKRDSGW DPAGSGHAQL PTAVALADGG VVLNGRVSAE
 951  RGGWLGGHVV AGTVLVPGAA LVEWVLRAGD EAGCPSLEEL TLQAPLVLPE
1001   SGGLQVQVVV GAADEQGGRR DVHVYSRSEQ DASAVWQCHA VGELGRASVA
1051   RPVRQAGQWP PAGAEPVEVG GFYEGVAAAG YEYGPAFRGL RAMWRHGDDL
```

1101 LAEVELPEEA GSPAGFGIHP ALLDAALHPL LAQRSRDGAG AGAHGGQVLL

1151 PFSWSGVSLW ASEATTVRVR LTGLGGGDDE TVSLTVTDPA GGPVVDVAEL

1201 RLRSTSARQV RGSAGPGADG LYELRWTPLP EPLPVPAPAN GRDVAADLSG

1251 CAVLGELVAE PGPGIDLEGC PCYPGVGALA DNASPPSMIL APVHSDTTGG

1301 DGLALTERVL RVIQDFLAAP SLEQKQTRLA FVTRGAADTG STTGGSAAPA

1351 EAVDPAVAAV WGLVRSAQSE NPGRFVLLDT DAPLDQASVA PLVDAVRSAV

1401 EADEPQVALR GGRLLVPRWA RAGEPVELAG PAGARAWRLV GGDSGTLEAV

1451 VAEACDDIVL RPLAPGQVRV AVHTAGVNFR DVLIALGMYP DPDALPGTEA

1501 AGVVTEVGPG VTRLSVGDRV MGMMDGAFGP WAVADARMLA PVPPGWGTRQ

1551 AAAAPAAFLT AWYGLVELAG LKAGERVLIH AATGGVGMAA VQIARHVGAE

1601 VFATASPGKH AVLEEMGIDA AHRASSRDLA FEDAFRQATD GRGVDVVLNS

1651 LTGELLDASL RLLGDGGRFV EMGKSDPRDP ELVALEHPGV SYEAFDLVAD

1701 AGPERLGLML DRLGELFAGG SLVPLPVTAW PLGRAREALR HMSQARHTGK

1751 LVLDVPAPLD PDGTVLVTGG TGTIGAAVAE HLARTGESKH LLIVSRSGPA

1801 AHGAEELVSR IAEFGAEATF VAADVSEPDA VAALIEGIDP AHPLTGVVHA

1851 AGVLDNALIG SQTTESLTRV WAAKAAAAQQ LHEATRESRL GLFVMFSSFA

1901 STMGTPGQAN YSAANAYCDA LAALRRAEGL AGLSVAWGLW EATSGLTGTL

1951 SAADRARIDR YGIRPTSAAR GCALLAAARA HGRPDLLAMD LDARVPAASD

2001 APVPAVLRTL AAAGAPATAR PTAAAAADGA TDWSGRLAGL TEEARLELLT

2051 ELVCTHAAGV LGHADAGAVQ VDAPFKELGF DSLTAVELRN RIAAATGLKL

2101 PAALVFDYPQ ARVLAAHLAE RLVPEGAGAM GGVSGAEGVR DAYGAGGPGG

2151 DMTAQVLLEV ARVEHTLSAA VPHGLDRAAV AARLEALLAR CTATTAATGA

2201 AGAAVEGDGD SDGDGAVDQL ETATAEQVLD FIDNELGV*

MonAIII, polyketide synthase multi-enzyme MONS3, housing extension modules 3 and 4 Length: 4133 amino acids

```
   1  MVSEEKLVDY LKRVSADLHA TRQRLREAEE RGQEPVAVVE AACRYPGGIR
  51  TPEDLWDLVA AGGNALGAFP DNRGWDLRRL FHPDPDHPGT TYAREGGFLH
 101  DADLFDPEFF GISPREAAVL DPQQRLLLEC AWEALERAGI DPRSLQGSRT
 151  GVYAGAALPG FGTPHIDPAA EGHLVTGSAP SVLSGRLAYT FGLEGPAVTI
 201  DTACSSSLVA VHLAAHALRQ RECDLALAGG VTVMTTPYVF TEFSRQRGLA
 251  ADGRCKPFAA AADGTAFSEG AGLLVLERLS DARRAGHRVL AVIRGSAVNQ
 301  DGASNGLTAP NGPAQQRVIR AALAGARLSP AEVDAVEAHG TGTRLGDPIE
 351  ADALLATYGQ ERHGGRPLWL GSVKSNIGHT QGAAGAAGLI KMVQALRHET
 401  LPATLYADEP TPHADWESGA VRLLSAPVAW PRGEHGEHTR RAGISSFGIS
 451  GTNAHLILEE APAADAEGAG GDGDGDGGGV RPVVRVGATG PREEQGQGQG
 501  QEQHQQQRQQ RQRSSMMPTP HLPWLLSARS PAALRAQADA LANHVAHADH
 551  SIADIGGTLL RRTLFEHRAV VLGTDRDERA AALAALAAGR AHPALTRAAG
 601  PARNGGTAFL FTGQGSQRPG MGRQLYDTFD VFAESLDETC ARLDPLLEQP
 651  LKPVLFAPAD TAQAAVLHGT GMTQAALFAL EVALYRQVTS FGIAPSHLTG
 701  HSVGEIAAAH VAGVFSLADA CTLVAARGRL MQALPAGGAM LAVQAAEDDV
 751  LPLLAGQEER LSLAAVNGPT AVVVSGEAAA VGEVEKALRG RGLKTKRLNV
 801  SHAFHSPLIE PMLDDFREVA RGLTFHAPTL PVVSNLTGRL ADAELMADAE
 851  YWVRHVRRPV RFHDGLRALS EQGVVRYLEL GPDPVLATMV QDGLPAPAEG
 901  EEPEPVVAAA LRSKHDEGRT LLGAVAALHT DGQPADLTAL FPADAGQVPL
 951  PTYRFQRRRY WRVAPDAAAP ARAAGLQETG HPLLPAVIRQ ADGGILLAGR
1001   LSLRTHPWLA DHTIAGGVPL PATAFVELAL LAGRHAACDT IDDLTLETPL
1051   LLDDTGTGVG AAVGAGADAL VDAIEVQLAL GAPDGSGRRA LTVHSRPADD
1101   AADDGDAADA ADAAGRGGPG GSGDLGDPGD PGDLGDGGGS RGWRRHATGI
```

1151 LSAGPAAEPA APDAAPWPPA DATALDVDAL YARLDAQGYS YGPAFRAVHA
1201 AWRHGDDLYA DVRLADEQRA EADAFALHPA LLDAALHAVD ELYRGSEGRG
1251 QEQGQGGQEP EQGRGDADAP VRLPFSFSDI RHHATGATRL WVRLSPQGDD
1301 RLRLSLTDGE GGQVATVDAL QLRLIPADRW RAARPTTAAP LYHLDWHELP
1351 LPEPAETDPA AHSWAVLGAH DAGLAPAAHY PDLAALKAAV EAGEPVPDIV
1401 FAPFPAQGTE TDVPAQVRAH ARHALELLRD WLTTEAFAAA RLVVLTTGAV
1451 TARPEDGPAD LATAPVWGLV RAAQAEQPDH VVLVDIDKDI DKDTDEETDQ
1501 ATDAGTASRH ALPAALAAAA AQAETQLALR AGTVLVPRLA VVPPRTDTPA
1551 LHATAPESTT DTVDSTGIAG AAESGGTVLI TGGTGGLGQA VARHLAAAHG
1601 ARHLLLVSRR GDAAEGVAEL RADLADDGVD VRVAACDITD RDALAGLLAD
1651 IPAAHPLTAV VHTAGVIDDS LITAMTPERL DAVLAPKADA AWHLHELTRD
1701 KDLSAFVLFS SGASVLGNGG QANYAAANTF LNTLAEHRRA AGLAATSVAW
1751 GLWESASGGM AARLGDADRA RIHRTGVTGL TDEQALALFD AALTAEHPTV
1801 LATRFDRAVL RGQAAAARTLQ PALRGLVRTP RPTASAGAIG STAATGSATD
1851 ENAPSSWAAR LARLSAADRD RALNELIREQ IATVLAHPSP DTIELGRAFQ
1901 ELGFDSLTAL ELRNRLSTAT GIRLPATLVF DHPSPTALVR HLHSHLPDEA
1951 QHTSPTAPGA SAEGTAATAT GIDDDPIAIV GMACRYPGGV TSPEQLWQLV
2001 ATGTDAIGPF PEDRGWDTAG LFDPDPDQVG HSYTREGGFL YDAARFDAGF
2051 FGISPREAAA TDPQQRLLLE TAWQAFEHAG IDPAALRGTP CGVITGIMYD
2101 DYGSRFLARK PDGFEGRIMT GSTPSVASGR VAYTFGLEGP AITVDTACSS
2151 SLVAMHLAAQ ALRQGECELA LAGGVTVMAT PNTFVEFSRQ RGLAPDGRCK
2201 PFAAAADGTG WGEGAGLVVL ERLSDARRKG HRVLALLRGS AVNQDGASNG
2251 MTAPNGPSQE RVIRTALAGA GRGPEDIDVV EAHGTGTTLG DPIEAQALLA
2301 TYGQGRPEDR PLWLGSVKSN IGHTQAAAGV AGVIKMVMAL RHEQLPTTLH

2351 ADEPTPHVQW DGGGVRLLTE PVPWSRGERT RRAGVSSFGI SGTNAHLILE

2401 EPPEEDLPEP VAAEPGGVVP WVVSGRTPDA LREQARRLGE FVVGAGDVSA

2451 AEVGWSLATT RSVFEHRAVV AGRDRDDLVA GMQALAAGET PTDVVSGAAA

2501 SSGAGPVLVF PGQGSQWVGM GAQLLDESPV FAARIAECEQ ALSAYVDWSL

2551 SDVLRGDGSE LSRVEVVQPV LWAVMVSLAA VWADYGVTPA AVVGHSQGEM

2601 AAACVAGALS LEDAARIVAV RSDALRQLQG HGDMASLGTG AEQAAELIGD

2651 RPGVVVAAVN GPSSTVISGP PEHVAAVVAE AEARGLRARV IDVGYASHGP

2701 QIDQLHDLLT EGLADIRPAN TDVAFYSTVT AERLTDTTAL DTDYWVTNLR

2751 QPVRFADTIE ALLADGYRLF IEASAHPVLG LGMEETIEQA DIPATVVPTL

2801 RRDHGDTTQL TRAAAHAFTA GADVDWRRWF PADPTPRTVD LPTYAFQHQH

2851 YWLEEPSGLT GDAADLGMVA AGHPLLGACV ELAESDSYLF TGRLSRRAPS

2901 WLAEHVVAGT VLVPGAALVE WVLRAGDEAG CPTIEELTLQ APLVLPESGG

2951 LQVQVVVGAT DEQSGRRDVH VYSRSEQDAS AVWVCHAVGV VSSEMPEAAA

3001 ELSGQWPPAG AEAVDVEDFY ARAAEAGYAY GPAFQGLRAL WRHGTELFAE

3051 VVLPEQAGGH DGFGIHPALL DAALHPLMLL DRPADGQMWL PFAWSGVSLN

3101 ADRATHVRVR LSPRGEAAER DLRVVIADAT GAPVLTVDAL TLRAADPGRL

3151 GAAARGGVDG LYTVDWTPLP LPQPLPLPRT DAGGSADWVI LSDNSSAALA

3201 DAVSSATAAG GGAPWALLAP VGGGSADDGL PVVRRTLSLV QEFLAAPELT

3251 ESRLVIVTRG AVATDADGDV AASAAAVWGL IRSAQSENPG RFVLLDVEEE

3301 HLHPDGGELP YAALRHAVEE LDEPQLALRS GKFLVPRMTP AAAPEELVPP

3351 VGTSGWRLGT SGTATLENLS VIDAPEAFAP LEPGQVRISV RAAGMNFRDV

3401 LIALGMYPDK GTFAGSEGAG HVTEVGPGVT HLSVGDRVMG LFEGAFAPLA

3451 VADARMVVPI PEGWSFQEAA AVPVVFLTAW YGLVDLGRLR AGESLLIHAG

3501 TGGVGMAATQ IARHLGAEVF ATASPAKHGV LDGMGIDAAH RASSRDLDFE

3551 ETLRAATGGR GMDVVLNSLA GEFTDASLRL LAEGGRMVDM GKTDKRDPDR
3601 VAAEHAGAWY RAFDLVPHAG PDRIGEMLAE LGELFASGAL APLPVQTWPL
3651 GRAREAFRFM SQAKHTGKLV LEIPPALDPD GTVLITGGTG VLAAAVAEHL
3701 VREWGVRHLL LAGRRGSEAP GSSELAEELT ELGAEVTFAA ADVSDPDAVA
3751 ELVGKTDPAH PLTGVIHAAG VLDDAVVTAQ TPESLARVWA AKATAAHLLH
3801 EATREARLGL FLVFSSAAAT LGSPGQANYA AANAYCDALV RQRRAEGLAG
3851 LSIGWGLWQT ASGMTGHLGE TDLARMKRTG FTPLTTEGGL ALLDAARAHG
3901 RPHVVAVDLD ARAVAAQPAP SRPALLRALA AGATPGARTA RRTAAAGSVA
3951 PAGGLADRLA GLPHPERRRL LLDLVRGNVA GVLGHSDHDA VRPDTSFKEL
4001 GFDSLTAVEL RNRLAAATGL KLPAALVFDY PESATLVDHL LERLSPDGAP
4051 PPVKDAADPV LNDLGRIESS LDALALDADA RSRVTRRLNT LLSKLNGAAT
4101 AGSPADVTDL DALDALDDVS DDEMFEFIDR EL*

MonAIV, polyketide synthase multi-enzyme MONS4, housing extension modules 5 and 6 Length: 4039 amino acids

1 MSSAEESSPD VSGTGVSGTG ESATGTSSTE AKLRQYLKRV TVDLGQARRR
51 LREVEERAQE PIAIVSMACR FPGDTRTPEA LWDLVAEGGD AIDDFPTNRG
101 WDLESLYHPD PDHPGTSYVR RGGFLYDAPA FDASFFGISP REALAMDPQQ
151 RVLMETAWQL LERAGIDPAS LKLSATGVYI GAGVLGFGGA QPDKTVEGHL
201 LTGSALSVLS GRISFTLGLE GPSVSVDTAC SSSLVSMHLA AQALRQGECD
251 LALAGGVTVM STPGAFTEFS RQGALSPDGR SKAFAASADG TGFSEGAGLL
301 LLERLSDARR NGHKVLAVIR GSAVNQDGAS NGLTAPNGPS QERVIRAALA
351 NAGLGAAEVD AVEAHGTGTK LGDPIEAGAL LATYGRDRDE DRPLWLGSVK
401 SNIGHPQGAA GVAGVIKMVM ALQRELLPAT LYVDEPTPHV DWSSGSVRLL
451 TEPVPWTRGE RPRRAGVSAF GMSGTNAHVI LEEAPPEEAA AAETPAEGTG
501 AVVPWVVSGR GEEALRAQAA QLAEHVRDDD QRPASPLEVG WSLATTRSVF

551 ENRAVVVGDD RDALLDGLRS LAAGEASPDV VSGAVGPTGP GPVMVFPGQG

601 GQWVGMGARL LDESPVFAAR IAECEQALSA YVDWSLTDVL RGDGSELARI

651 DVVQPVLWAV MVALAAVWAD QGIEPAAVVG HSQGEIAAAC VVGAISLDEA

701 ARIVAVRSVL LRQLSGRGGM ASLGMGQEQA ADLIDGHPGV VVAAVNGPSS

751 TVISGPPEGI AAVVADAQER GLRARAVASD VAGHGPQLDA ILDQLTEGLA

801 GIRPAATDVA FYSTVTAGHL TDTTELDTAY WVRNVRRTVR FADTIDALLA

851 DGYRLFIEVS PHPVLNLALE GLIERAAVPA TVVPTLRRDH GDTTQLARAA

901 AHAFAAGADV DWRRWFPADP APRTVDLPTY AFQRQDFWPA PAGGRSGDPA

951 GLGLAASGHP LLGASVGLAS GDVHLLSGRV SRQSAAWLDD HVVAGQALVP

1001 GAAQVEWVLR AGDDAGCSAL EELTLQTPLV LPDTGGLRIQ VVVEAADAHG

1051 RRDVRLFSRP DDDDAFASTH PWTCHATGVL APAPTDGTNG TRDAADTLDG

1101 AWPPADAEPV PADDLYAQAD RTGYGYGPAF RGVRALWRHG KDVLAEVTLP

1151 KEAGDPDGFG IHPALLDAVL QPAALLLPPT DAEQVWLPFA WNDVALHAVR

1201 ATTVRVRLTP LGERIDQGLR ITVADAVGAP VLTVRDLRSR PTDTGRLAAA

1251 ATRDRHGLFD LEWIAPENAA ENAAGPARDA SEGWVTLGED AASLADLLAS

1301 VEAGAPAPQL VAAPVEPDRT DDGLALATHV LDLVQTWLAS PLHDSRLVLV

1351 TRGAVTDADV DVAAAAVWGL VRSAQSEHPG RFTLIDLGPD DTLAAAMQAA

1401 HLEEPQLAVH GGEIRVPRLV RATTDPTAPN GTPEADRTAD PSEGLHRNGT

1451 VLITGGTGVL GRLVAEHLVT EWGVRHLLLA SRRGDQAPGS AELRARLSEL

1501 GASVEIAPAD VGDAEAVAAL IASVDPAHPL TGVIHAAGVL DDAVITAQTP

1551 ESLARVWATK ATAARHLHEA TRETPLDFFV VFSSAAASLG SPGQANYAAA

1601 NAYCDALVQH RRAQGLAGLS IAWGLWQATS GMTGQLSETD LARMKRTGFA

1651 ALTDEGGLAL LDAARAHDRA YVVAADLDPR AVTDGLSPLL RALTAPATRR

1701 RVASEGLADG ALATRLAGLD ADGRLRLLTD VVREYVAAVL GHGSAARVGV

1751 DIAFKDLGFD SLTAVELRNR LSAACDVRLP ATLIFDHPTP QALATHLVDR

1801 LAGSTSATTT VNATAPAAAH VAAGADVDAD TDDPVAIVAM TCRFPGGVAS

1851 PDDLWDLLDA RKDAMGAFPT DRGWDLERLF HPDPDHPGTS YTDQGGFLPD

1901 AGDFDAAFFG INPREALAMD PQQRLLLEAS WEVLERAGID PTTLKGTPTG

1951 TYVGLMYHDY AKSFPTADAQ LEGYSYLAST GSMVSGRVAY TLGLEGPAVT

2001 VDTACSSSLV SIHLATQALR HGECDLALAG GVTVMADPDM FAGFSRQRGL

2051 SPDGRCKAYA AAADGVGFSE GVGVLLLERL SDARRHGRRV LGVVRGSAVN

2101 QDGASNGLTA PNGPSQERVI RQALASGGLS SVDVDVVEGH GTGTTLGDPI

2151 EAQALLATYG QGRPEDRPLW LGSVKSNIGH TQAAAGVAGV IKMVMAMRHG

2201 VVPASLHVDV PSPHVEWDSG AVRLAVESVP WPQVEGRPRR AGVSSFGASG

2251 TNAHVIVESV PDGLEEDSVS VGGEALETET DGRLVPWVVS ARSPQALRDQ

2301 ALRLRDFASD ASFRAPLADV GWSLLKTRAL HEHRAVVVGA ERAELIAALE

2351 ALATGEPHAA LVGPACSQAR VGGDDVVWLF SGQGSQLVGM GAGLYERFPV

2401 FAAAFDEVCG LLEGPLGVEA GGLREVVFRG PRERLDHTVW AQAGLFALQV

2451 GLARLWESVG VRPDVVLGHS IGEIAAAHVA GVFDLADACR VVGARARLMG

2501 GLPEGGAMCA VQATPAELAA DVDGSAVSVA AVNTPDSTVI SGPSDEVDRI

2551 AGVWRERGRK TKALSVSHAF HSALMEPMLA EFTEAIRGVK FRQPSIPLMS

2601 NVSGERAGEE ITDPEYWARH VRNAVLFQPA IAQVADSAGV FVELGPAPVL

2651 TTAAQHTLDE SDSQESVLVA SLAGERPEES AFVEAMARLH TAGVAVDWSV

2701 LFAGDRVPGL VELPTYAFQR ERFWLSGRSG GGDAATLGLV AAGHPLLGAA

2751 VEFADRGGCL LTGRLSRSGV SWLADHVVAG AVLVPGAALV EWALRAGDEV

2801 GCVTVEELML QAPLVVPEAS GLRVQVVVEE AGEDGRRGVQ IYSRPDADAV

2851 GGDDSWICHA TGVLSPESAR LDTELGGVWP PAGAEPLDVD GFYAQAGEAG

2901 YGYGPAFRGL RAVWRHGQDL LAEVVLPEAA GAHDGYGIHP ALLDATLHPL

```
2951  LAARFMDGSE DDQLYVPFGW AGVSLRAVGA TTVRVRLRPV GESVDQGLSV
3001  TVTDATGGPV LSVDSLQTRP VKPSQLAAAQ QPDVRGLFTV EWTPLPQTDA
3051  DGEADWVVLS DGVGRLADVV SAAGGEAPWA VVAPVDASVG DGREGLDGRL
3101  VVERVLSLVQ EFLALPELAE SRLLVVTRGA VATGVDGDGD VDASAAAVWG
3151  LVRSAQSENP GRFILLDVDG DGDDQGPDLN GRHLPHATLR HAAEELDEPQ
3201  LALREGTLYV PRLTQARQSA ELVVPPGEPA WRLRMVHDGS LDALAAVACP
3251  EALEPLAPGQ VRIAVHAAGI NFRDVLVALG MVPAYGAMGG EGAGVVTEVG
3301  PEVTHVSVGD RVMGVFEGAF GPVVIAEARM VTPVPQGWDM REAAGIPAAF
3351  LTAWYGLVEL AGLKAGERVL VHAATGGVGM AAVQIARHVG AEVFATASPG
3401  KHAVLEEMGI DAAHRASSRD LAFEGTFREA TGGRGMDVVL NSLAGEFIDA
3451  SLRLLGDGGR FLEMGKTDVR AAEEVAAEHA DVSYTAYDLV GDAGPDRISN
3501  MLDKLVELFA SERLKPLPVR SWPLDKAQEA FRFMSQAKHT GKLVLEIPPA
3551  LDPEGTVLVT GGTGALGQVV AEHLVREWGV RHLLLASRRG PEAPGSDELA
3601  SKLTGLGAEV TIVAADVSDP ASVVELVGKT DPSHPLTGVV HAAGVLEDGV
3651  VTAQTPEGLA RVWAAKAAAA ANLHEATREM RLGLFVVFSS AAATLGSPGQ
3701  ANYAAANAYC DALMQHRRAV GQVGLSVGWG LWEAPDAKPG VAADAKASAA
3751  TVGKASALSD GTNGSAPQDT TGTAPQGMTG GLTDTDVARM ARIGVKGMSN
3801  AHGLALFDAA HRHGRPHLVG FNLDLRTLAT HPLHTRPALL RGLATPTAGG
3851  ASRPTATAGG QPADLAGRLA ALSPSDRHHT LVRLIREQAA TVLGHHPDSL
3901  TTGSTFKELG FDSLTAVELR NRLSAATGLR LPAGLVFDHP DADILAEHLG
3951  AQLAPDGDTP AGAEATDPVL RDLAKLENAL SSTLVEHLDA DAVTARLEAL
4001  LSNWKAASAA PGSGSTKEQL QVATTDQVLD FIDKELGV*
```

MonAV, polyketide synthase multi-enzyme MONS5, housing extension modules 7 and 8 Length: 4107 amino acids

```
   1  MASEEELVDY LKRVAAELHD TRQRLREVED RRQEPVAVVG MACRFPGGIE
  51  TPEGLWELVA AGDDAIEPFP TDRGWDLEGI YHPDPDHPGT CYVREGGFLA
 101  APDRFDSDFF GFSPREALAS SPQLRLLLET SWEALERAGI NPASLKGSPT
 151  GVYVGAATTG NQTQGDPGGK ATEGYAGTAP SVLSGRLSFT LGLEGPAVTV
 201  ETACSSSLVA MHLAANALRQ GECDLALAGG VTVMSTPEVF TGFSRQRGLA
 251  PDGRCKPFAA AADGTGWGEG AGLILLERLS DARRKGHKVL AVIRGSAINQ
 301  DGASNGFTAP NGPSQRRVIR QALSSAHLST SEIDVVEAHG TGTRLGDPIE
 351  AEALIATYGK EREDDRPLWL GSVKSNIGHT QAAAGVAGVI KMVMALQREL
 401  LPATLNVDEP TPHVQWEGGG VRLLTEPVPW SRGERPRRAG ISSFGISGTN
 451  AHVVLEEAPP EEDVPGPVAA EPEGVVPWVV SARTEEALSE QARRLGEFVA
 501  DTDPSTADVG WSLTTSRAIL EHRAVVVGRD RDALTAGLAA LAAGEESADV
 551  VAGVAGDVGP GPVLVFPGQG SQWVGMGAQL LDESPVFAAR IAECEQALSA
 601  YVDWSLSAVL RGDGSELSRV EVVQPVLWAV MVSLAAVWAD YGVTPAAVIG
 651  HSQGEMAAAC VAGALSLEDA ARVVAVRSDA LRQLMGQGDM ASLGASSEQA
 701  AELIGDRPGV CIAAVNGPSS TVISGPPEHV AAVVADAEER GLRARVIDVG
 751  YASHGPQIDQ LHDLLTDRLA DIRPATTDVA FYSTVTAERL TDTTALDTDY
 801  WVTNLRQPVR FADTIDALLA DGYRLFIEAS AHPVLGLGME ETIEQADIPA
 851  TVVPTLRRDH GDTTQLTRAA AHAFTAGATV DWRRWFPADP TPRTIDLPTY
 901  AFQRRSYWLP VDGVGDVRSA GLRRVEHSLL PAALGLADGA LVLTGRLAAS
 951  GGGGGWLADH AVAGTTLVPG AALVEWALRA ADEAGCPSLE ELTLQAPLVL
1001  PGSGGLQVQV VVGPADGQGG RREVRVFSRV DSDDEAAGQD EGWSCHATGV
1051  LSPEPGAVPD GLSGQWPPTG AEPLEISDLY EQAASAGYEY GPSFRGLRSV
1101  WRHGHNLLAE VELPEQAGAH DDFGIHPVLL DAALHPALLL DQNAPGEEQE
1151  PAQPALRLPF VWNGVSLWAT GAATVRVRLA PHGGGETDDS AGLRVTVADA
```

1201 TGAPVLSVDS LALRPADPEL LRTAGRAGSG TNGLFTVEWT ALPPADVADH
1251 AAGDGWAVLG QDVPDWAGAD MPRHPDMASL SAALDEGTQA PAAVFVETTA
1301 TSHATPNTAA DVTLDASGRA VAERTLHLLR DWLAEPRLAE TRLVLITHHA
1351 VTTPADDDVN AAPLDVPAAA LWGLIRSAQA EHPDRFVLLD TDAKANTDPG
1401 PDTSTDHSTA SGTYRTVIAR ALATGEPQLA VRAGELLAPR LARAATPTPE
1451 TPTPETQPDT GSGSEAGAGS GSGPGATLDP DGTVLIAGGT GMMGGLVAEH
1501 LVRAWSVRHL LLVSRQGPDA PDARDLADRL VGLGATVRIV AADLTDGRAT
1551 ADLVASVDPA HPLTGVIHAA GVLDDAVVTA QTSDQLARVW AAKASVAANL
1601 DAATSELPLG LFLMFSSAAG VLGNAGQAGY AAANAFVDAL VGRRRATGLP
1651 GLSIAWGLWA RGSAMTRHLD DADLARLRAG GVKPLLDEQG LALLDAARAT
1701 AAHTSLVVAA GIDVRGLNRD DVPAILRDLA GRTRRRAAAD STVDQAALER
1751 RLTGLDEAER RAVVTDVVRE CVAAVLGHRS AADVRTEANF KDLGFDSLTA
1801 VQLRNRLSAA SGLRLPATLA FDHPTPQALA AYLGTRLSGR TATPVAPVAP
1851 SAAATDEPVA IVAMACKYPG GATSPEGLWD LVAEGVDAVG AFPTGRGWDL
1901 ERLFHPDPDH PGTSYADEGA FLPDAGDFDA AFFGINPREA LAMDPQQRLL
1951 LEASWEVLER AGIDPTTLKG TPTGTYVGVM YHDYAAGLAQ DAQLEGYSML
2001 AGSGSVVSGR VAYTLGLEGP AVTVDTACSS SLVSIHLAAQ ALRQGECTLA
2051 LAGGVTVMAT PEVFTGFSRQ RGLAPDGRCK PFAAAADGTG WGEGVGVLLL
2101 ERLSDARRHG RRVLGVVRGS AVNQDGASNG LTAPNGPSQE RVIRQALASG
2151 GLSSVDVDVV EGHGTGTTLG DPIEAQALLA TYGQGRPVDR PLWLGSVKSN
2201 IGHTQAAAGV AGVIKMVMAM RHGVVPASLH VDVPSPHVEW DSGAVRLAVE
2251 SVPWPEVEGR PRRAGVSSFG ASGTNAHVIV ESVPDGLGED SVSVSGEAPE
2301 TETDGRLVPW VVSARSPQAL RDQALRLRDA VAADSTVSVQ DVGWSLLKTR
2351 ALFEQRAVVV GRERAELLSG LAVLAAGEEH PAVTRSREDG VAASGAVVWL

2401 FSGQGSQLVG MGAGLYERFP VFAAAFDEVC GLLEGPLGVE AGGLREVVFR
2451 GPRERLDHTM WAQAGLFALQ VGLARLWESV GVRPDVVLGH SIGEIAAAHV
2501 AGVFDLADAC RVVGARARLM GGLPEGGAMC AVQATPAELA ADVDDSGVSV
2551 AAVNTPDSTV ISGPSGEVDR IAGVWRERGR KTKALSVSHA FHSALMEPML
2601 AEFTEAIREV KFTRPKVSLI SNVSGLEAGE EIASPEYWAR HVRQTVLFQP
2651 GIAQVASTAG VFVELGPGPV LTTAAQHTLD DVTDRHGPEP VLVSSLAGER
2701 PEESAFVEAM ARLHTAGVAV DWSVLFAGDR VPGLVELPTY AFQRERFWLS
2751 GRSGGGDAAT LGLVAAGHPL LGAAVEFADR GGCLLTGRLS RSGVSWLADH
2801 VVAGAVLVPG AALVEWALRA GDEVGCVTVE ELMLQAPLVV PEASGLRVQV
2851 VVEEAGEDGR RGVQIYSRPD ADAVSGDDSW ICHATGTLTP QHTDAPNDGL
2901 AGAWPAAGAV PVDLAGFYER VADAGYAYGP GFQGLRAVWR HGQDLLAEVV
2951 LPEAAGAHDG YGIHPALLDA TLHPALLLDW PGEVQDDDGK VWLPFTWNQV
3001 SLRAAGAATV RVRLSPGEHD EAEREVQVLV ADATGTDVLS VGSVTLRPAD
3051 IRQLQAVPGH DDGLFSVDWT PLPLSRTDVS QTDADGDADW VVLSDGVGSL
3101 ADVVSAAGGE APWAVVAPVG ASAGGGLAGF DRREGLDGRL VVERVLSLVQ
3151 EFLAAPELAE SRLLVLTRGA VATGGDGDGD VDASAAAVWG LVRSAQSENP
3201 GRFILLDVDM DVDVDVDMDV DVDVDVDVDV DGDGNGSDLD PDLNGRRLPH
3251 ATLRHAAEEL DEPQLALRDG QLLVPRLVRA TGGGLVVAPT DRAWRLDKGS
3301 AETLESVAPV AYPGVMEPLG PGQVRLGIHA AGINFRDVLV SLGMVPGQVG
3351 LGGEGAGVVT ETGPDVTHLS VGDRVMGVLH GSFGPTAVAD TRMVAPVPQG
3401 WDMRQAAAMP VAYLTAWYGL VELAGLKAGE RVLIHAATGG VGMAAVQIAR
3451 HLGAEVFATA SAAKHVVLEE MGIDAAHRAS SRDLAFEDTF RQATDGRGMD
3501 VVLNSLTGEF IDASLRLLGD GGRFLEMGKT DVRTPEEVAA EYPGVTYTVY
3551 DLVTDAGPDR IAVMMSELGE RFASGALDPL PVRSWPLDKA REAFRFMSQA

3601 KHTGKLVLDV PAPLDPDGTV LITGGTGALG QVVAEHLVRE WGVRHLLLAS

3651 RRGLDAPGSG ELADRLSDLG AEVTVAAADV SDPASVVELV GKTDPSHPLT

3701 GVVHAAGVLE DGIVTAQTPE GLARVWAAKA AAAANLHEAT REMRLGLFVV

3751 FSSAAATLGS PGQANYAAAN AYCDALMQRR RAAGQVGLSV GWGLWEAPDA

3801 KPGVAADAKP DVAADAKTGV AADGTPQGMT GTLSGTDVAR MARIGVKAMT

3851 SAHGLALLDA AHRHGRPHLV AVDLDTRVLA HKPAPALPAL LRAFAGDQGG

3901 QGGGRGGGRG GGPARPAAAT TRQNVDWAAK LSVLTAEEQH RTLLDLVRTH

3951 AAAVLGHAGT DAVRADAAFQ DLGFDSLTAV ELRNRLSAST GLRLPATFIF

4001 RHPTPSAIAD ELRAQLAPAG ADPAAPLFGE LDKLETVITG HAHDESTRTR

4051 LAARLQNLLW RLDDTSARSD HAAGASDADG DAVENRDLES ASDDELFELI

4101 DRELPS*

MonAVI, polyketide synthase multi-enzyme MONS6, housing extension module 9 Length: 1701 amino acids

1 MPGTNDMPGT EDKLRHYLKR VTADLGQTRQ RLRDVEERQR EPIAIVAMAC

51 RYPGGVASPE QLWDLVASRG DAIEEFPADR GWDVAGLYHP DPDHPGTTYV

101 REAGFLRDAA RFDADFFGIN PREALAADPQ QRVLLEVSWE LFERAGIDPA

151 TLKDTLTGVY AGVSSQDHMS GSRVPPEVEG YATTGTLSSV ISGRIAYTFG

201 LEGPAVTLDT ACSASLVAIH LACQALRQGD CGLAVAGGVT VLSTPTAFVE

251 FSRQRGLAPD GRCKPFAEAA DGTGFSEGVG LILLERLSDA RRNGHQVLGV

301 VRGSAVNQDG ASNGLTAPND VAQERVIRQA LTNARVTPDA VDAVEAHGTG

351 TTLGDPIEGN ALLATYGKDR PADRPLWLGS VKSNIGHTQA AAGVAGVIKM

401 VMAMRHGELP ASLHIDRPTP HVDWEGGGVR LLTDPVPWPR ADRPRRAGVS

451 SFGISGTNAH LIVEQAPAPP DTADDAPEGA ATPGASDGLV VPWVVSARSP

501 QALRDQALRL RDFAGDASRA PLTDVGWSLL RSRALFEQRA VVAGRERAEL

551 LAGLAALAAG EEHPAVTRSR EEAAVAASGD VVWLFSGQGS QLVGMGAGLY

601 ERFPVFAAAF DEVCGLLEGE LGVGSGGLRE VVFWGPRERL DHTVWAQAGL

651 FALQVGLARL WESVGVRPDV VLGHSIGEIA AAHVAGVFDL ADACRVVGAR

701 ARLMGGLPEG GAMCAVQATP AELAADVDGS SVSVAAVNTP DSTVISGPSG

751 EVDRIAGVWR ERGRKTKALS VSHAFHSALM EPMLGEFTEA IRGVKFRQPS

801 IPLMSNVSGE RAGEEITSPE YWARHVRQTV LFQPGVAQVA AEARAFVELG

851 PGPVLTAAAQ HTLDHITEPE GPEPVVTASL HPDRPDDVAF AHAMADLHVA

901 GISVDWSAYF PDDPAPRTVD LPTYAFQGRR FWLADIAAPE AVSSTDGEEA

951 GFWAAVEGAD FQALCDTLHL KDDEHRAALE TVFPALSAWR RERRERSIVD

1001 AWRYRVDWRR VELPTPVPGA GTGPDADTGL GAWLIVAPTH GSGTWPQACA

1051 RALEEAGAPV RIVEAGPHAD RADMADLVQA WRASCADDTT QLGGVLSLLA

1101 LAEAPATSSD TTSHTSTSCG TGSLASHGLT GTLTLLHGLL DAGVEAPLWC

1151 ATRGAVSCGD ADPLVSPSQA PVWGLGRVAA LEHPELWGGL VDLPADPESL

1201 DASALYAVLR GDGGEDQVAL RRGAVLGRRL VPDATPDVAP GSSPDVSGGA

1251 AHADATSGEW QPHGAVLVTG GVGHLADQVV RWLAASGAEH VVLLDTGPAN

1301 SRGPGRNDDL AAEAAEHGTE LTVLRSLSEL TDVSVRPIRT VIHTSLPGEL

1351 APLAEVTPDA LGAAVSAAAR LSELPGIGSV ETVLFFSSVT ASLGSREHGA

1401 YAAANAYLDA LAQRAGADAA SPRTVSVGWG IWDLPDDGDV ARGAAGLSRR

1451 QGLPPLEPQL ALGALRAALD GGKGHTLVAD IEWERFAPLF TLARPTRLLD

1501 GIPAAQRVLD ASSESAEASE NASALRRELT ALPVRERTGA LLDLVRKQVA

1551 AVLRYEPGQD VAPEKAFKDL GFDSLVVVEL RNRLRAATGL RLPATLVYDY

1601 PTPRTLAAHL LDRVLPDGGA AELPVAAHLD DLEAALTDLP ADDPRRKGLV

1651 RRLQTLLWKQ PDAMGAAGPA DEEEQAAPED LSTASADDMF ALIDREWGTR

1701 *

MonH, probable regulatory protein Length: 981 amino acids

```
  1  VSGVERGVGS AGPVEQGDGL AGLVERAEAL AALRGAFDGS PGTGGSLVVL
 51  SGAVGTGKTA LLRAWADRIG ADADALVLTA TACRAERDLP LGVLEQLVRS
101  PGLPPASAER ALAWWDEEAS ATPGKTDANG TSANGTDANG TGAGQTGAGQ
151  AGVGQTGVGG EPVLAASALR GLCEVLRDLL AERPVVVAVD DAHHADAASL
201  QCLLSVVRRL RSARLHVLFT EYAHQKAQNA LLSSEFLHEP ALRRIRLEPL
251  SKAGVEALLA RHLDERTAQD LTPVVHGMSA GHPLLVRALA EDHRAAGGAG
301  EAYGRAVLSF LYRHETPVTQ VARAIAALGA HAGPGQVGRL LDVDAASVER
351  AVRQLTVAEV LHEGRLCHPA FAAAVLDGMP PEERRALHGR VADLLHEEGA
401  PATEVAAHLV AADRSDAPWA VPVFQEAAQL ALDEDQVETG VDYLRAAHQR
451  CRGAAQRAAV VGALADAEWR LDPAKVLRHL PDPAAMAPQT DPAALAPHTD
501  PAPTAAPTAA PTPTPIPTTP PLPTHLLWHG RVEEGLDAIG TLTGPGPNPA
551  GAPPMNPADL DTPWLWGAYL YPGHVKERLG SGALSPQRST PPAVTPELQG
601  AGTLMNDLLH GGERDATEAA ERALNRYRLG PRTIAVQTAA LAALTYRDRP
651  HRAAAWCDGL VAQADERNSP TWRALFTAWR ALLHLRQGDP AAAEQRAETA
701  LALLGSKGWG AAIGLPLAAA VQAKAALGDV DGAAALLERP VPQAVFQTRT
751  GLHYLAARGR YHLATGCHYA ALCDFYACGT RMSSWGVDLP ALEPWRLGAA
801  EAYLALGEGL LARQLVDGQL PLPTPDDGRT WGMTLRLRAA TSPAPARAEL
851  LDEAVAVLRE SGDTFELARA VADQAVAVRE GGEAERARLL ARKAELLARR
901  WGSAPAPATV PEPPERPGPA TPDAELTSAE RRVAELAAEG FTNREISRKL
951  CVTVSTVEQH LTRIYRKLDV RRLDLQAALG *
```

MonCI, flavin-dependent epoxidase Length: 496 amino acids

```
  1  VTTTRPAHAV VLGASMAGTL AAHVLARHVD AVTVVERDAL PEEPQHRKGV
 51  PQARHAHLLW SNGARLIEEM LPGTTDRLLA AGARRLGFPE DLVTLTGQGW
101  QHRFPATQFA LVASRPLLDL TVRQQALGAD NITVRQRTEA VELTGSGGGS
```

151 GGRVTGVVVR DLDSGRQEQL EADLVIDATG RGSRLKQWLA ALGVPALEED

201 VVDAGVAYAT RLFKAPPGAT THFPAVNIAA DDRVREPGRF GVVYPIEGGR

251 WLATLSCTRG AQLPTHEDEF IPFAENLNHP ILADLLRDAE PLTPVFGSRS

301 GANRRLYPER LEQWPDGLLV IGDSLTAFNP IYGHGMSSAA RCATTIDREF

351 ERSVQEGTGS ARAGTRALQK AIGAAVDDPW ILAATKDIDY VNCRVSATDP

401 RLIGVDTEQR LRFAEAITAA SIRSPKASEI VTDVMSLNAP QAELGSNRFL

451 MAMRADERLP ELTAPPFLPE ELAVVGLDAA TISPTPTPTP TAAVRS

MonBII, carbon-carbon double bond isomerase Length: 141 amino acids

1 MPDEAARKQM AVDYAERINA GDIEGVLDLF TDDIVFEDPV GRPPMVGKDD

51 LRRHLELAVS CGTHEVPDPP MTSMDDRFVV TPTTVTVQRP RPMTFRIVGI

101 VELDEHGLGR RVQAFWGVTD VTMDDPAGPA DTTHPEGIRA *

MonBI, carbon-carbon double bond isomerase Length: 144 amino acids

1 MNEFARKKRA LEHSRRINAG DLDAIIDLYA PDAVLEDPVG LPPVTGHDAL

51 RAHYEPLLAA HLREEAAEPV AGQDATHALI QISSVMDYLP VGPLYAERGW

101 LKAPDAPGTA RIHRTAMLVI RMDASGLIRH LKSYWGTSDL TVLG

MonAVIII, polyketide synthase multi-enzyme MONS8, housing extension modules 11 and 12 Length: 3754 amino acids

1 MSNEEKLLDH LKWVTAELRQ ARQRLHDKES TEPVAIVGMA CRYPGGARSA

51 EDLWELVRDG GDAVAGFPDD RGWDLESLYH PDPEHPATSY VRDGAFLYDA

101 GHFDAEFFGI SPREATAMDP QQRLLLETAW EAIEHAGMNP HALKGSDTGV

151 FTGVSAHDYL TLISQTASDV EGYIGTGNLG SVVSGRISYT VGLEGPAVTV

201 DTACSSSLVA IHLASQALRQ GECSLALAGG STVMATPGSF TEFSRQRGLA

251 PDGRCKPFAA AADGTGWGEG AGVVALELLS EARRRGHKVL AVIRGSATNQ

301 DGTSNGLAAP NGPSQERVIR AALANARLSA EDIDAVEAHG TGTTLGDPIE

351 AQALIATYGQ GRPEDRPLWL GSVKSNIGHT QAAAGVAGVI KMVMAMRNGL

401 LPTSLHIDAP SPHVQWEQGS VRLLSEPVDW PAERTRRAGI SAFGISGTNA

451 HLILEEAPPE EDAPGPVAAE PGGVVPWVVS GRTPDALREQ ARRLGEFAAG

501 LADASVSEVG WSLATTRALF DQRAVVVGRD LAQAGASLEA LAAGEASADV

551 VAGVAGDVGP GPVLVFPGQG SQWVGMGAQL LDESPVFAAR IAECEQALSA

601 HVDWSLSDVL RGDGSELSRV EVVQPVLWAV MVSLAAVWAD YGITPAAVIG

651 HSQGEMAAAC VAGALSLEDA ARIVAVRSDA LRQLQGHGDM ASLSTGAEQA

701 AELIGDRPGV VVAAVNGPSS TVISGPPEHV AAVVADAEAQ GLRARVIDVR

751 YASHGPQIDQ LHDLLTDRLA DIQPTTTDVA FYSTVTAERL DDTTALDTAY

801 WVTNLRQPVR FADTIEALLA DGYRLFIEAS PHPVLNLGIQ ETIEQQAGAA

851 GTAVTIPTLR RDHGDTTQLT RAAAHAFTAG APVDWRRWFP ADPTPRTVDL

901 PTYAFQHKHY WVEPPAAVAA VGGGHDPVEA RVWQAIEDLD IDALAGSLEI

951 EGQAESVGAL ESALPVLSAW RRRHREQSTV DSWRYQVTWK HLPDVPAPEL

1001 SGAWLLLVPA AHADHPAVLA TAQTLTAHGG EVRRHVVDAR AMERTELAQE

1051 LRVLMDGAAF AGVVNLLALD EEPHPEHSAV PAGLAATTAL VQALADNGAD

1101 IAVRTLTQGA VSTSAGDALT HPVQAQVWGL GRVAALEYPR LWGGLVDLPA

1151 RIDHQTLARL AAALVPQDED QISIRPSGVH ARRLAHAPAN TVGSGLGWRP

1201 DGTTLITGGT GGIGAVLARW LARAGAPHLL LTSRRGPDAP GAQELAAELT

1251 ELGAAVTVTA CDVGDREQVR RLIDDVPAEH PLTAVIHAAG VPNYIGLGDV

1301 SGAELDEVLR PKALAAHHLH ELTRELPLSA FVMFSSGAGV WGSGQQGAYG

1351 AANHFLDALA EHRRAEGLPA TSIAWGPWAE AGMAADQAAL TFFSRFGLHP

1401 LSPELCVKAL QQALDAGETT LTVANFDWAQ FTSTFTAQRP SPLLADLPEN

1451 RRASAPAAQQ EDATEASSLQ QELTEAKPAQ QRQLLLQHVR SQAAATLGHS

1501 DVDAVPATKP FQELGFDSLT AVELRNRLNK STGLTLPTTV VFDHPTPDAL

1551 TDVLRAELSG DAAASADPVR AAGASRGAAD DEPIAIVGMA CRYPGDVRSA

1601 EELWDLVAAG KDAMGAFPDD RGWDLETLYD PDPESRGTSY VREGGFLYDA

1651 GDFDAGFFGI SPREAVAMDP QQRLLLETAW EAIERAGLDR ETLKGSDAGV

1701 FTGLTIFDYL ALVGEQPTEV EGYIGTGNLG CVASGRVSYV LGLEGPAMTI

1751 DTGCSSSLVA IHQAAHALRQ GECSLALAGG ATVMATPGSF VEFSLQRGLA

1801 KDGRCKPFAA AADGTGWAEG VGLVVLERLS EARRNGHNVL AVIRGSAINQ

1851 DGTSNGLTAP NGQAQQRVIR QALANARLSA EDVDAVEAHG TGTMLGDPIE

1901 ASALVATYGK ERPADRPLWL GSIKSNIGHA QASAGVAGVI KMVMALRNEQ

1951 LPASLHIDAP TPHVDWDGSG VRLLSEPVSW PRGERPRRAG VSAFGISGTN

2001 AHLILEQAPD APEPVTAPAE DAAAPAGVVP WVVSARGEEA LRAQARLLAD

2051 RATADPRLAS PLDVGWSLVK TRSVFENRAV VVGKDRQTLL AGLRSLAAGE

2101 PSPDVVEGAV QGASGAGPVL VFPGQGSQWV GMGAQLLDES PVFAARIAEC

2151 ERALSAHVDW SLSAVLRGDG SELSRVEVVQ PVLWAVMVSL ASVWADYGIT

2201 PAAVIGHSQG EMAAACVAGA LSLEDAARIV AVRSDALRQL MGQGDMASLG

2251 AGSEQVAELI GDRPGVCVAA VNGPSSTVIS GPPEHVAAVV ADAEARGLRA

2301 RVIDVGYASH GPQIDQLHDL LTERLADIRP TTTDVAFYST VTAERLDDTT

2351 TLDTDYWVTN LRQPVRFADT IEALLADGYR LFIEASPHPV LNLGMEETIE

2401 RADMPATVVP TLRRDHGDAA QLTRAAAQAF GAGAEVDWTG WFPAVPLPRV

2451 VDLPTYAFQR ERFWLEGRRG LAGDPAGLGL ASAGHPLLGA AVELADGGSH

2501 LLTGRISPRD QAWLAEHRVM DTVLLPGSAF VELALQAAVR AGCAELAELT

2551 LHTPLAFGDE GAGAVDVQVV VGSVAEDGRR PVTVHSRPTG EGEEAVWTRH

2601 AAGVVAPPGP DAGDASFGGT WPPPGATPVG EQDPYGELAS YGYDFGPGSQ

2651 GLVSAWRLGD DLFAEVALPE AESGRADRYQ VHPVLLDATL HALILDAVTS

2701 SADTDQVLLP FSWSGLRVHA PGAEKLRVRI ARTAPDQLAL TAVDGGGGGE

2751 PVLTLESLTV RPVAAHQIAG ARAADRDALF RLVWMEVAAR AEETGGGAPR

2801 AAVLAPVESG PMGGTSAGAL ADALSDALAA GPVWDTFGAL RDGVAAGGEA

2851 PDVVLAVCAA PGAGAGAVAD ADGRGGDPAG YARLATVSLL SLLKEWVDDP

2901 AFAATRLVVV TRGAVAARPG ETAGDLAGAS LWGLVRSAQA ENPGRLTLLD

2951 VDGLESSPAT LTGVLASGEP ELALRDGRAY VPRLVRDDAS VRLVPPVGSL

3001 TWRLARCQEA GGGQQLSLVD APEAGRALEP HEVRVAVRAA APGPLTAGQV

3051 EGAGVVTEVG GEVGSVAVGD RVMGLFDAVG PVAVTDAALL MPVPAGWSWA

3101 QAAGSLGAYV SAYHVLADVV APRGGETLLV GEETGSVGRA VLRLALAGRW

3151 RVEAVDGAST ADDSGAERAA DVTLRHEGAL VVHRAGGRPD EGQAVVPPEP

3201 GRVREILAEL TELTELAEIT ESAEPGLPAE RGDSRALTPL DITVWDIRQA

3251 PAAMAAPPSA GTTVFSLPPA FDPEGTVLVT GGTGALGSLT ARHLVERYGA

3301 RHLLLSSRRG ADAPGALELA ADLSALGARV TFAACDPGDR DEAAALLAAV

3351 PSDHPLTAVF HCAGTVNDAV VQNLTAEQVE EVMRVKADAA WHLHELTRDA

3401 DLSAFVLYSS VAGLLGGPGQ GSYTAANAFL DALARHRHDG GAAATSLAWG

3451 YWELASGMSG RLTDADRARH ARAGVVGLGA DEGLALLDAA WAGGLPLYAP

3501 VRLDLARMRR QAQSHPAPAL LRDLVRGGSK SGGGAVSAGA AALLKSLGAM

3551 SDPEREEALL DLVCTHIAAV LGYDAATPVN ATQGLRELGF DSLTAVELRN

3601 RLSAATGLKL PATFVFDHPN PAELAAQLRQ ELAPRAADPL ADVLAEFERI

3651 EDSLLSVSSK DGSARAELAG RLRATLARLD APQDTAGEVA VATRTRIQDA

3701 SADEIFAFID RDLGRDGASG QGNGQPTGQG NGHGNGNGNG NGNGHGQAVE

3751 GQR*

MonAVII, polyketide synthase multi-enzyme MONS7, housing extension module 10 Length: 1642 amino acids

1 MAHTEEKLLE YLKRVTADLR QTERRLQDVE SAGHEPVAVI GMACRLPGGV

51 RSPEEFWELV STGGDAVAPL PGNRNWDLDS LYDPDPESTG TSYVREGGFV

```
101  YDAGDFDPTF FGIGPTEAAA MAPQQRLALE TAWEAIERAG IDPLSLRSSD
151  TSTFIGCDGL DYALGASEVP EGTAGYFTIG NSGSVTSGRV AYTLGLEGPA
201  VTVDTACSSS LVSLHLATQA LRTQECSLAL AGGTYVMSSP APLIGFSELR
251  GLAPDGRCKP FSASSDGMGM AEGTGVVLLE RLSDARRKGH KVLAVIRGSA
301  INQDGASNGL TAPNGPAQER VIRAALANAR LAPEDIDAVE AHGTGTTLGD
351  PIEAGALISA YGRERPEDRP LWVGAVKSNI GHTQIAAGVA GVIKMVLALR
401  HDLLPAILHV DAPSPHVEWD GSGLRLLTDP VKWPRGERPR RAGVSSFGFS
451  GTNAHLILEE APPEEEDVPG SVAEEPGGVV PWVVSGRTPD ALRAQARRLG
501  EFAAGPADAS AADVGWSLTT TRSVFEHRAV VVGRDRDALT AGLGALAAGE
551  ASAGVVAGVA GDVGPGPVLV FPGQGSQWVG MGAQLLDESP VFAARIAECE
601  RALSAYVDWS LSAVLRGDGS ELSRVEVVQP VLWAVMVSLA AVWADYGVTP
651  AAVIGHSQGE MAAACVAGAL SLEDAARIVA VRSDALRRLQ GHGDMASLST
701  GAEQAAELIG DRPGVVVAAV NGPSSTVISG PPEHVAAVVA DAEARGLRAR
751  VIDVGYASHG PQIDQLHDLL TERLADIRPA NTDVAFYSTV TAERLTDTTA
801  LDTDYWVTNL RQPVRFADTI EALLADGYRL FIEASAHPVL GLGMEETIEQ
851  ADIPATVVPT LRRDHGDTTQ LTRAAAHAFT AGAPVDWRRW FPADPTPRTV
901  DLPTYAFQHQ HYWLERSASA SGAVSGEQSA AEAQLWHAVE ELDLGLLAET
951  LGSEEGSEEA VRALEPALPV LKGWRRRHQD QATIDSWRYR VTWKQRSDGP
1001  APELGGDWLL FVPADKAEHP AVRATAEALS EHGAAAVRLH PVETGRAGRQ
1051  ELAAVDTAGL AGIVNLLALD EEPHPEHPAV PAGLAATTAL LQALGDNGTT
1101  APLHTVTQGA VSTGATDPLT HPLQAHVWGL GRVAALEHPR LWAGLVDLPA
1151  RIDRHTLPRL AAALLPQDDE DQTAVRPTGI HHRRLTHAVG SIQNPVHSEA
1201  TWRPRGTTLI TGGTGGIGAV LARWLARQGA PRLHLTSRRG PDAPGARELA
1251  AELDGLGTAV TITACDVSDP RQLSGLIDDM PAEHPLTAVI HAAGMTDLTA
```

1301 IGDLTTARLG EVLGSKSDAA WNLHELTRDL DLSAFVMFSS GAGVWGSGQQ

1351 GAYGAANHFL DALAEHRRAQ GLPATSIAWG PWAEAGMSAD PESLTYFKRF

1401 GLLPIAPDLC VKALHQAVDA GDATLTVANF DWAKFTPTFT AQRPSPFLDD

1451 LPENQREAEQ TGTAAETSAF REELAKTPAS QRLGFLVQQV RTYAAATLGR

1501 TVEDIPAAKP FQELGFDSLT AVQLRNQLNT TTGLSLPATV IFDHPTPEAL

1551 ATHLRGQLGD GAEVAGEGDV LAALDKWDTA FGAAEVDEAA RRRIVGRLQV

1601 LVSKWSPAQD GPEGTDSAHA DLEAASADDI FDLISSEFGK S*

MonD, cytochrome P450 hydroxylase Length: 431 amino acids

1 VGLTVGPDNA KRGIVPITDS KPAATFPDLV DPSFWARPHA ERVALFEEMR

51 GLPRPAFIRQ NMPGVPWTFG YHALVKYADI VEVSRRPQDF SSNGATTIIG

101 LPPELDEYYG SMINMDNPEH SRLRRIVSRS FGRNMIPEFE AVATRTARRI

151 IDELIARGPG DFIRPVAAEM PIAVLSDMMG IPAEDHDFLF DRSNTIVGPL

201 DPDYVPDRAD SERAVIEASR ELGDYIAGLR AERLAAPGND LITKLVQVQA

251 DGEQLTRQEL VSFFILLVIA GMETTRNAIS HALVLLTEHP EQKQLLLSDF

301 DTHAPNAVEE ILRVSTPINW MRRVATRDCD MNGHRFRRGD RIFLFYWSGN

351 RDESVFPDPY RFDITRGTNA HVTFGAVGPH VCLGAHLARM EITVLYRELL

401 AALPQIHAVG QPRRLDSSFI EGIKHLHCAF *

MonRI, probable activator protein Length: 268 amino acids

1 VRYEMLGPLR IKDGNDYATI NAQKVEIVLT VLLIRADRVV SLEQLMREIW

51 GEDLPRRATA GLHVYISQLR KFLKVPGSAG NPVETRAPGY VLHKRDDDQI

101 DAQIFPELVD VGRSLLREKR FDEAASCFGQ ALALWRGPIL GQGGNGPGTN

151 GPIIDGFSTW LTEIRLECQE MLVECQLQLG RHREAVGMLY ALTAENPMCE

201 AFYRQLMLAL YRSERQADAL KVYQSVRKTL NDELGLEPGR PLQELQRAIL

251 AGDMHLMSPP PLALSGR*

MonAX, thioesterase Length: 278 amino acids

```
  1  LSAFLAKGKI LSAFPPPDMS DPWIRRFRPR PEAVVRLVCF PHAGGSASYY
 51  HPLAQSPTLP TDSEVLAVQY PGRQDRRRER LLDDIGELAD LITDALGPFD
101  DRPLAFFGHS MGAVLAYEVA QRLRERTGKQ PCRLFVSGRR APSRFRRGTV
151  HLLDDTELAA ELRRAGGTDP RFLDDEELLA EIIPVVRNDY RAVELYRWNP
201  SPPLSCPITA LVGDRDPQAP LDEVEAWQQH TEGPFDLKVF AGGHFYLNTH
251  QQGVTEVISK ALADSAQQRA TARGNAR*
```

ORF29, a homologue of CapK involved in cell wall biosynthesis Length: 428 amino acids

```
  1  LADLVAHARS ASPYYRELYH GLPERIEDPT LLPVTDKKQL MDHFDDWPTD
 51  RDITFEKVRA FTDDPELIGR RFLGRYLVAT TSGTSGRRGL FVLDDRYMNV
101  SSAVSSRVLA SWLGPLGIAR AVVHGGRFAQ LVATEGHYVG FAGYSRLRQD
151  GEARSKLVRA FSVHEPMSRL VAELNEYRPA FVIGYASTIM LFTAEQEAGR
201  LHIDPVLVEP AGETMTESDT DRIAAAFGAK VRTMYSATEC TYLSHGCAEG
251  WYHVNDDWAV LEPVDADHRP TPPGEFSHTT LISNLANRVQ PFLRYDLGDS
301  VMLRPDPCPC GTPSPAIRVQ GRSGDILTFP SGRGDDVSLA PLAFSSLFDR
351  MPGVELFQIE QTAPSTLRVR VVQAPGADAD HVWQRAHDGL THLLADNKLD
401  NVTVERGEEP PRQASGGKYR TIIPLAA*
```

LipB, lipase B Length: 338 amino acids

```
  1  VKVPVEVTVR LSSWLGGLVA AVLAATVLPA SAASAADVSS PPLEIPAAEL
 51  AKALHCGTEL GDLRDAGDKP TVLFVPGTGL KGEENYAWNY MAELKKKGYQ
101  SCWVDSPGRG LRDMQESVEY VVYATRAIQE ATGRKVDLVG HSQGGLLTAW
151  ALRFWPDLPG KVDDMVTLGS PFQGTRLASP CRPIAEVAGC PASVLQFARD
201  SNWSKALGAD GTPMPAGPSY TTIYSYADES VVADGEAPSL PGAHRIGVQD
```

```
251  ICPGRPWPTH IAMVVDQVSY DLVADAIEHP GPADTSRIDR AHCAKPVMPL
301  NSQEAVDALP GLLNFPIELL IHSQPWVDEE PPLRPYAR
```

ORF31, putative ion pump Length: 309 amino acids

```
  1  MGHDHGPSAG AAGGTLSGTY RKRLLWTIGI SGSITVIQVV GALLSGSLAL
 51  LADAAHSLTD AVGVSLALGA ITLAQRAPTP RRTFGFCRVE IFSAVLNALL
101  LVVIFAWVLW SAIGRFSEPV EVKGGLMFVV ALGGLAANLV GLWLLRDAKE
151  KSLNLRGAYL EVLGDALGSV AVIVGGLVIL LTGWQAADPI ASIVIGLLIV
201  PRAYGLLRDS LHVLLEATPQ DVDLGEVRRH LLEERGVVAV HDLHGWTVTS
251  GMPVLTAHVV VTEEALASGY GELLGRLQRC VGGHFDVAHS TIQLEPEGHV
301  EEDGALHT*
```

ORF32, hypothetical membrane protein Length: 364 amino acids

```
  1  MTRALTLHDW IVAGIAVVAG VVAGLLLRAL LRWLGERASK TRWSGDDVIV
 51  DALRTLVPCA AITAGLAAAA GALPLTPRTG RNVTMTLTAL LILAATLTAA
101  RIVTGLVKAV AQSRSGVAGS ATIFVNITRV VVLAMGFLIV LQTLGISIAP
151  LLTALGVGGL AVALALQDTL ANLFAGVHIL AAKTVQPGDY IQLSSGEEGY
201  VVDINWRNTT VRQLSNNLVI IPNAKLAGTN MTNYSRPEQE LSIMVQVGVS
251  YDSDLEQVEK VTTEVVDEVM AEITGAVPDH EAAIRFHTFG DSRISFTVIL
301  GVGEFSDQYR IKHEFIKRLH QRYRAEGIRV PAPVRTVRVQ QGELPPPLGI
351  PHQRDTSTQA RLH*
```

AmtA, glycine amidinotransferase (partial coding sequence) Length: 131 amino acids

```
  1  MSPVNSHNEW DPLEEIIVGR LEGATIPSSH PVVACNIPTW AARLQGLAAG
 51  FEYPQRLIEP AQQELDQFIA LLQSLDVTVR RPAAVDHKHR FGTPDWQSRG
101  FCNSCPRDSM LVVGDEIIET PMAWPCRCFE T
```

CLAIMS:

1. A DNA sequence which is (a) at least part of the sequence set out in the appended sequence listing; or (b) a variant of a sequence (a) which encodes a polypeptide which is at least 80%, preferably at least 90%, identical with the corresponding peptide as set out in table II; provided that it is not a sequence encoding all or part of the polypeptide consisting of amino acids 1-920 encoded by *mon AI* as set out in table II.

2. A DNA sequence according to claim 1 comprising the complete monensin gene cluster or a variant thereof.

3. A DNA sequence encoding at least part of at least one polypeptide which is necessary for the biosynthesis of monensin, and which is encoded by DNA included in the appended sequence listing or an allele, mutation or other variant thereof; provided that said polypeptide is not all or part of amino acids 1-920 encoded by *mon AI* as set out in table II.

4. A DNA sequence according to claim 3 which comprises at least part of one or more of the following genes: *mon BI, mon BII, mon CI, mon CII, mon H, mon RI, mon RII, mon T, mon AIX* and *mon AX*.

5. A DNA sequence according to claim 4 comprising all of the genes listed therein or an allele, mutation or other variant thereof.

6. A DNA sequence according to claim 3 encoding at least part of one or more of the polypeptides set out below, said polypeptide having the amino acid sequence as set out in the appended sequence data or being a variant thereof having the specified activity:

peptide	activity
mon CII	epoxyhydrolase/cyclase
mon E	S-adenosylmethionine-dependent methyltransferase
mon T	monensin resistance gene
mon RII	repressor protein
mon AIX	thioesterase
mon AI	polyketide synthase multienzyme
mon AII	polyketide synthase multienzyme
mon AIII	polyketide synthase multienzyme
mon AIV	polyketide synthase multienzyme
mon AV	polyketide synthase multienzyme
mon AVI	polyketide synthase multienzyme
mon AVII	polyketide synthase multienzyme
mon AVIII	polyketide synthase multienzyme
mon H	regulatory protein
mon CI	flavin-dependent epoxidase
mon BII	carbon-carbon double bond isomerase

mon BI carbon-carbon double bond isomerase

mon D cytochrome P450 hydroxylase

mon RI activator protein

mon AX thioesterase

7. A DNA sequence according to claim 6 encoding a single enzyme activity of a multienzyme encoded by any of *mon AI-mon AVIII* or a variant or part thereof.

8. A DNA sequence according to any preceding claim encoding any one or more of the domains as set out in Table I or a variant or part thereof.

9. A DNA sequence according to any preceding claim which has a length of at least 30, preferably at least 60, bases.

10. A recombinant cloning or expression vector comprising a DNA sequence according to any preceding claim.

11. A transformant host cell which has been transformed to contain a DNA sequence according to any of claims 1-9 and which is capable of expressing a corresponding polypeptide.

12. A hybridisation probe which is a DNA sequence according to any of claims 1-9.

13. Use of a probe according to claim 12 to detect a PKS cluster, optionally followed by isolation of the detected cluster.

14. Use of a probe according to claim 12 which encodes at least part of a polypeptide having a known function to detect genes encoding polypeptides having analogous function.

15. Use according to claim 14 wherein the polypeptide of known function is AT of module 5 or the regulatory protein encoded by *mon RI*.

16. A hybridization probe comprising a polynucleotide which binds specifically to a region of the monensin gene cluster selected from *mon BI, mon BII, mon CI, mon CII, mon H, mon RI, mon RII, mon T, mon AIX* and *mon AX*.

17. Use of a probe according to claim 16 in a method of detecting the presence of a gene cluster which governs the synthesis of a polyether, and optionally isolating a gene cluster detected thereby.

18. Use of a probe according to claim 12 which comprise a polynucleotide which binds specifically to a gene responsible for levels of activity of the monensin gene cluster, in a method of detecting an analogous gene in a gene cluster for biosynthesis of another polyketide, optionally followed by a step of manipulating the gene detected thereby to alter the level of expression of said other polyketide.

19. Use according to claim 18 wherein the gene is a regulatory gene, resistance gene or thioesterase gene.

20. Use of the *mon RI* gene or variant and a monensin promoter to control expression of a heterologous gene in *S. cinnamonensis*.

21. Use of a portion of the monensin gene cluster encoding a polypeptide having chain terminating activity, preferably comprising at least one of *mon AIX* and *mon AX* or a mutant, allele or other variant thereof encoding a polypeptide having chain terminating activity, to effect chain release of a peptide other than monensin.

22. Use of a portion of the monensin gene cluster encoding a polypeptide having carbon-carbon double bond isomerase activity, preferably comprising at least one of

mon BI and *mon BII* or a mutant, allele or other variant thereof having isomerase activity to provide a desired stereochemical outcome in the synthesis of a polyketide other than monensin.

23. A polypeptide encoded by a portion of the monensin gene cluster, preferably comprising at least one of *mon BI* and *mon BII* or a mutant, allele or other variant thereof, having carbon-carbon double bond isomerase activity, or at least one of *mon AIX* and *mon AX* or a mutant, allele or other variant thereof having chain terminating activity.

24. An epoxidase enzyme encoded by *mon CI* or a derivative or variant thereof having epoxidase activity.

25. A cyclase enzyme encoded by *mon CII* or a derivative or variant thereof having cyclase activity.

26. Use of a portion of the monensin gene cluster encoding a peptide having epoxidase or cyclase activity, preferably comprising *mon CI* or *mon CII* or a mutant, allele or other variant thereof encoding a polypeptide having epoxidase or cyclase activity to provide a said activity in the biosynthesis of a polypeptide other than monensin.

27. A process for producing a polyketide containing a desired starter unit comprising providing a PKS gene having a loading module and a plurality of extension modules, wherein the loading module includes a KS_q domain derived from a KS domain of a monensin extension module.

28. A process according to claim 27 wherein the KS_q domain is derived from KS of module 5 of monensin.

29. A process according to claim 27 or claim 28 wherein the starter unit also includes an AT_q domain derived from an AT domain which is naturally associated with the KS domain.

30. A DNA sequence comprising DNA encoding at least one PKS loading module and a plurality of PKS extension modules, and which can be expressed to produce a polyketide; wherein at least one of said modules or at least one domain thereof is a monensin module or domain or a variant thereof and is contiguous to a further one of said modules or a domain to which it is not naturally contiguous; provided that the sequence is not an ery loading module, the first and second extension modules of the ery PKS and the ery chain-terminating thioesterase in which the DNA encoding AT of the first extension module has been substituted by DNA encoding an ethyl malonyl-CoA

AT from the monensin gene cluster.

31. A DNA sequence according to claim 30 wherein said further module or domain is also a monensin module or domain or variant thereof.

32. A DNA sequence according to claim 30 wherein said further module or domain is a module or domain of a PKS of a polyketide other than monensin or a variant thereof.

33. A DNA sequence according to claim 30, 31 or 32 wherein said loading module is adapted to load a starter unit other than a starter unit normally received by the adjacent extension module.

34. A DNA sequence according to claim 33 wherein said loading module is derived from a monensin extension module or variant thereof.

35. A polyketide synthase encoded by the DNA sequence of any of claims 30-34.

36. A polyketide compound as produced by a synthase according to claim 35.

37. A vector containing a DNA sequence of any of claims 30-34.

38. A transformant cell transformed to contain a DNA sequence of any of claims 30-34.

39. A method of producing *S. cinnamonensis* capable of enhanced levels of production of monensin comprising engineering it to overexpress the *mon RI* gene.

40. A method according to claim 39 wherein said engineering comprises introducing at least one additional copy of the *mon RI* gene as shown in the appended sequence data or a variant thereof.

41. *S. cinnamonensis* containing multiple copies of the *mon RI* gene as shown in the appended sequence data and/or variant(s) thereof.

42. A method of producing monensin comprising culturing the organism of claim 41 and/or an organism produced by the method of claim 39 or claim 40.

43. A process for expressing a gene heterologous to *S. cinnamonensis* comprising transforming *S. cinnamonensis* with DNA encoding a heterologous gene and expressing said

gene under control of the activator gene *mon RI* or actII/orf4.

44. A process according to claim 43 wherein said heterologous gene is a PKS gene.

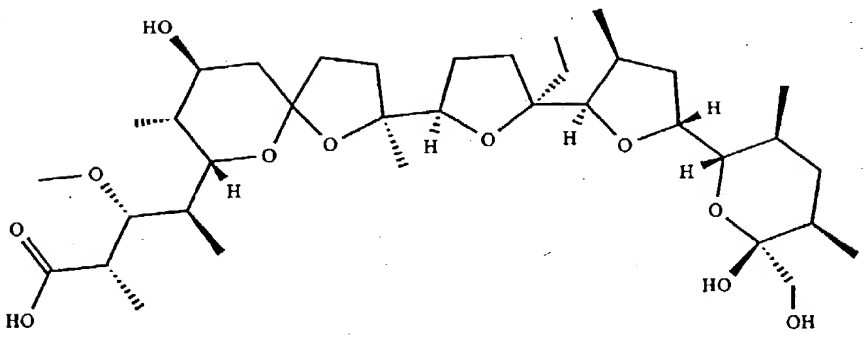
45. 13-Propyl erythromycin A.

POLYKETIDES AND THEIR SYNTHESIS

ABSTRACT

The complete sequence of the gene cluster for the monensin type I polyketide synthase, from *S. cinnamonensis*, is provided. Thus variant polyketides containing monensin-derived elements can be genetically engineered. Furthermore there are novel features, e.g. a regulatory protein *mon RI*, which are of wide utility.

1/4



monensin A : R = ethyl
monensin B : R = methyl

Fig 1

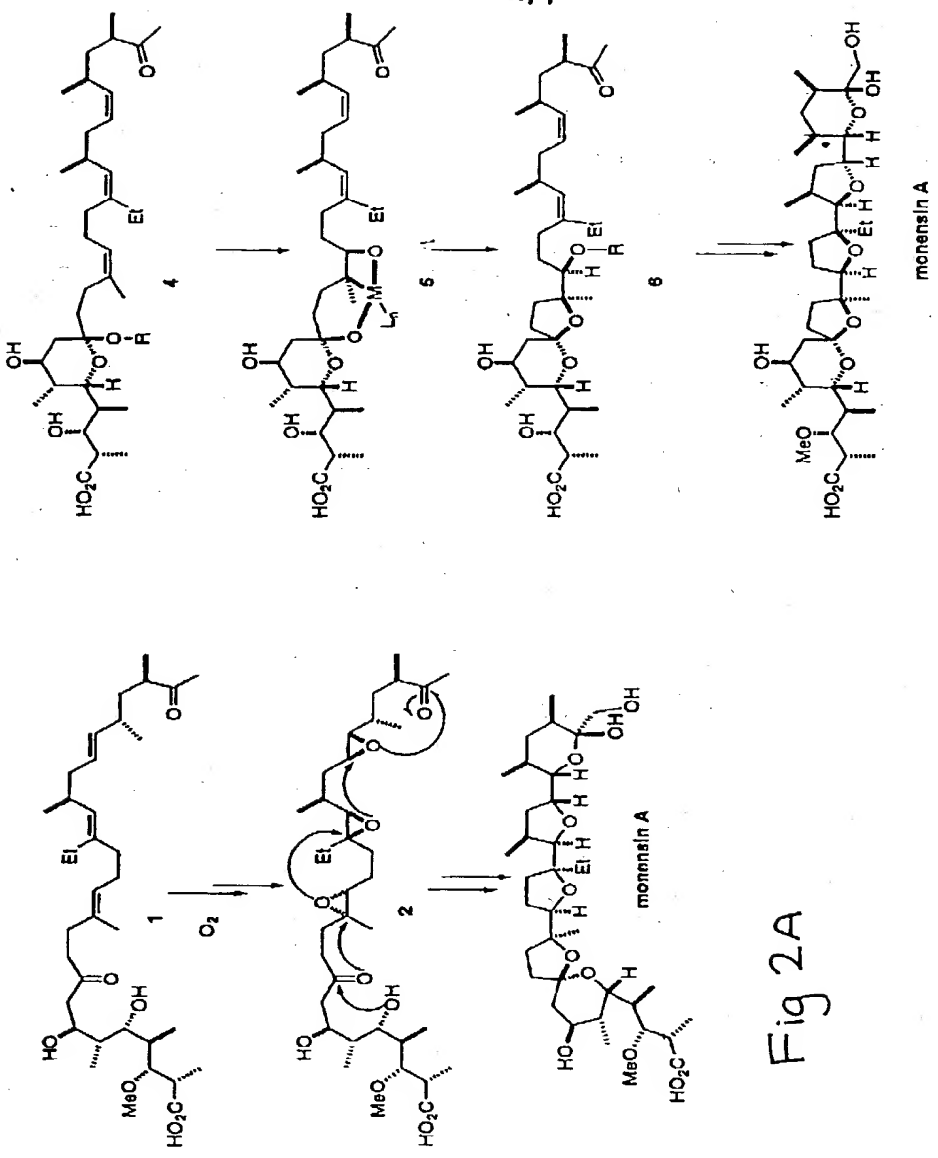


Fig 2A

Fig 2B

Figure 2. Proposed mechanisms for monensin biosynthesis.

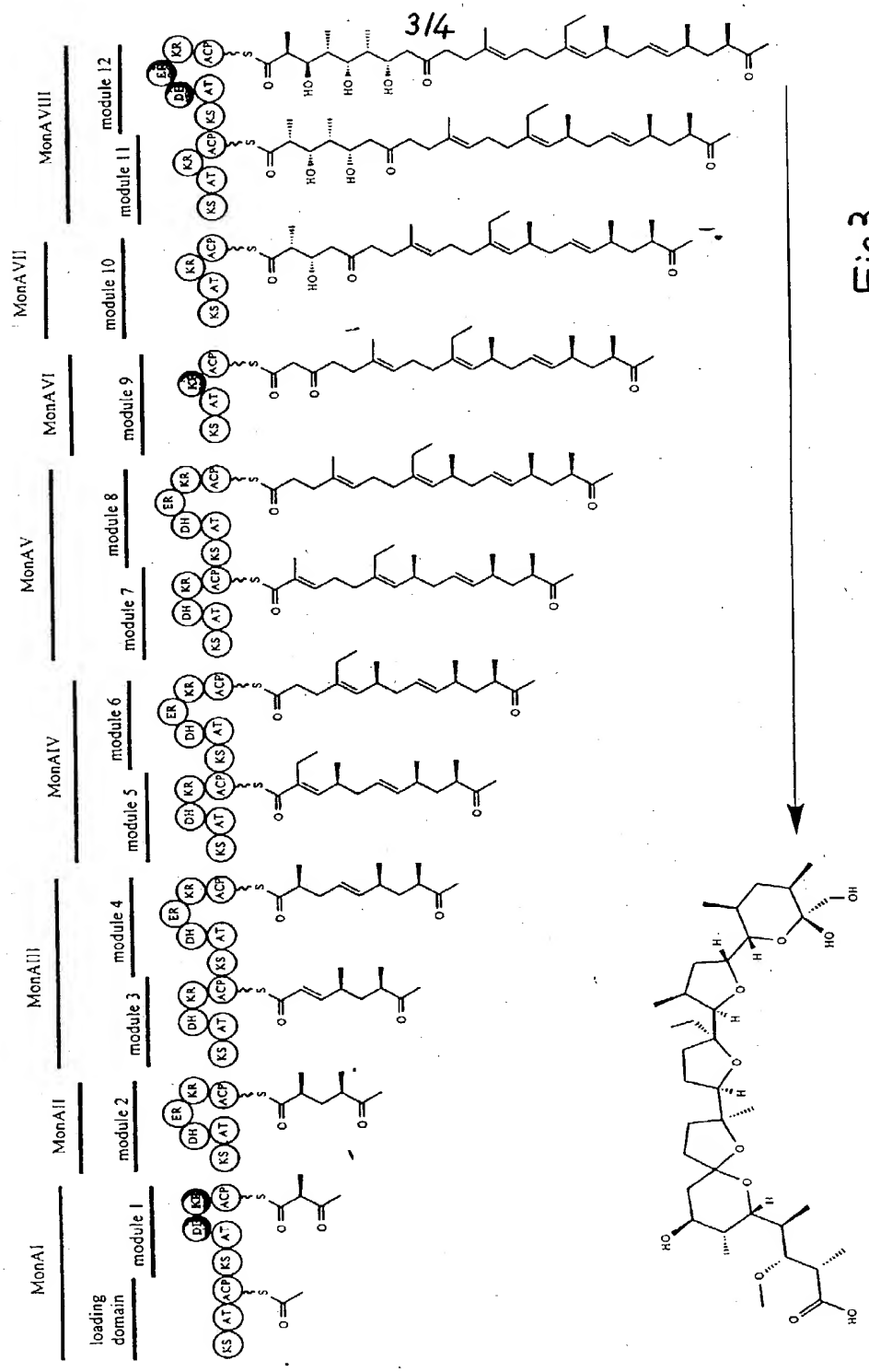


Fig3

Proposed organisation of the monensin PKS

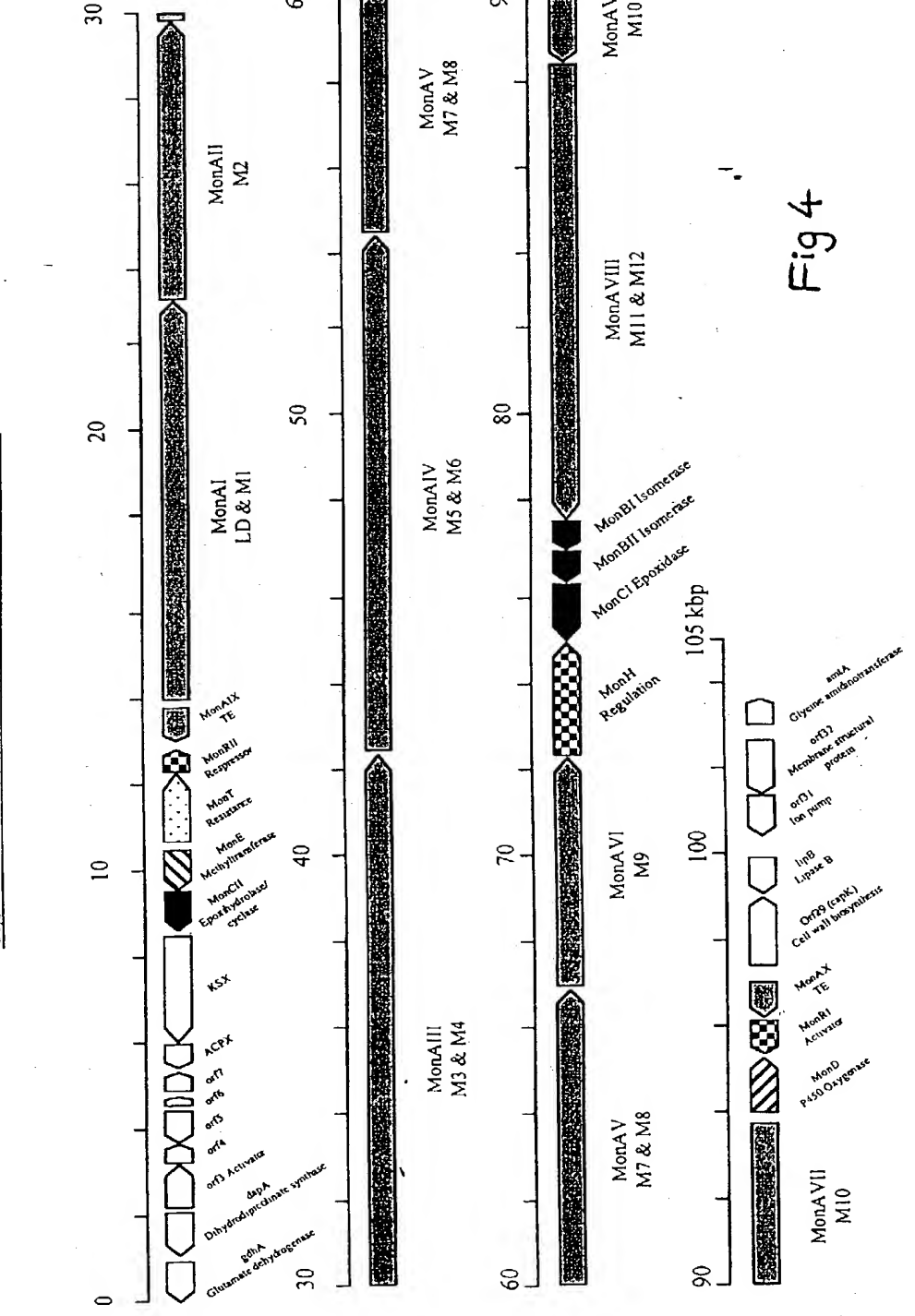

Organisation of the Monensin Biosynthetic Gene Cluster
0
10
20
30
40
50
60
70
80
90
100
105 kbp
gdhA Glutamate dehydrogenase
dapA Dihydrodipicolinate synthase
orf3 Activator
orf4
orf5
orf6
orf7
ACPX
KSX
MonCII Epoxyhydrolase/ cyclase
MonE Methyltransferase
MonT Resistance
MonRII Repressor
MonAIX TE
MonAI LD & M1
MonAII M2
MonAIII M3 & M4
MonAIV M5 & M6
MonAV M7 & M8
MonAVI M9
MonAVII M10
MonAVIII M11 & M12
MonH Regulation
MonCI Epoxidase
MonBII Isomerase
MonBI Isomerase
MonD P450 Oxygenase
MonRI Activator
MonAX TE
Orf29 (cspK) Cell wall biosynthesis
lipB Lipase B
orf31 Ion pump
orf32 Membrane structural protein
amiA Glycine amidinotransferase

Fig 4

SEQUENCE LISTING

```
   1  GATCAGCGCG GTGGCGTCGT CGGCGTCCAG CTCGTTCTGC GTGGCGGACG
  51  GCAGCGCGAT GTCGGCAGGC ACCTCCCAGA CCCGGCGGCC CGGCACGAAG
 101  CGGGCCGAGG CGCCGCGGCG CTGGGCGTAG GTGTCCACGC GGGCGCGTTC
 151  GACCTCCTTG ACCTGCTTGA GGAGGTCCAG GTCGATGCCC TTCTCGTCGA
 201  CGACGTAACC GGAGGAGTCC GAACACGTCA CGGCGTTGGC GCCCAGGGCG
 251  GCGAGCTTCT GGATGGTGTA GATGGCGACG TTCCCGGAGC CGGACACGAC
 301  CGCCGTCCGG CCTTCGAGGG TCTCGCCGCG CTCACGCAGC ATCGCCGCCG
 351  CGAAGAGGAC GTTGCCGTAG CCGGTCGCCT CCGGACGGAT CAGGGAGCCG
 401  CCCCAGTTGC GGCCCTTGCC GGTGAGGACG CCCGCCTCCC AGCGGTTGGT
 451  GATGCGCCGG TACTGACCGA ACAGATAGCC GATCTCCCGG CCGCCGACGC
 501  CGATGTCGCC CGCGGGCACG TCCGTGTGTT CGCCGATGTG CCGGTACAGC
 551  TCCGTCATGA ACGACTGGCA GAAACGCATG ACTTCCGCGT CGCTGCGGCC
 601  GCGCGGGTCG AAGTCGCTGC CGCCCTTGCC GCCGCCGATG CCGAGGCCCG
 651  TCAGCGCGTT CTTGAAGATC TGCTCGAAGC CCAGGAACTT GATGACGCCG
 701  AGGTTCACCG ACGGGTGGAA GCGCAGGCCG CCCTTGTACG GGCCGAGGGC
 751  GCTGTTGAAC TCCACCCGGA AGCCGCGGTT GACCCGCACG CGACCGTGGT
 801  CGTCCTGCCA CGGCACCCGG AAGACGATCT GGCGCTCCGG TTCGCACAGG
 851  CGCTCGATCA GGCCGGCTTC GGCGTACTCG GGGCGAGCCG CGATGACCGG
 901  CGCCAGGGTC TCGAGGACCT CGCGGGCGGC CTGGTGGAAC TCCGGCTGGG
 951  CCGGGTTGCG GTGTTCGATC TCGGTGAGCA GCTGGGAGAG TGCTGTCTTC
1001  TGCGAGAGAG CTGTCTTCGT GTCGGGTCGC GTGGTCAAAG GAGCCCTTTC
1051  TGGCACGGCC GGCGTAGGCG CTCGGCGCCG TTGCCGTGCG CAGGGAGACG
1101  CTCGAGCCGC AAGTATGACG CGCATGTAAA CACAGCGACC AGCCCCCCGG
1151  TCCAGGGAGT GACCACCATG CGAGACCGGG CCACCGGTAG GGCCACCGGT
1201  CCGGCCTGCG GACCCCGTGT CACTTCCGGC TCGCGGCCAG GGGTGCCGCC
```

```
1251  CGGCGGACCG AATCGGCGGA GGCGGCCAGC AGTGGCATGC GGACGGCCGG
1301  GCTGGGAATG CGGTTCTGGG CGTGCAGCAC TCCCTTGATC ACCGTCGGGT
1351  TCGGTTCGGT GAAGAGGGCG GCGGAAAGGC GGGCGAGGTC GGCTCCGAGA
1401  GCGCGGGCGG GTGCGGCGGA GCCGCGTCGC CACAGCCCGA TCATCTCGGC
1451  GTAGTCGGCG GTACGCAGAT TGGCCGACGC CACGATTCCG CCGTGGGCGC
1501  CCGCAGCGAC CAGCGGCGAG AGGACGATGT CGTCACCGCC GAGCACGGCG
1551  AAGCCGGGCA GGGGCGAGTC GAGCAACTCC ATGGTGGTCG GGTCGATCGA
1601  GCCGGTCGCG TGCTTGATGC CGACGACCTC CGGCAGGCGG CCGAGTGCGG
1651  TGATCGTGCC CGCGCCGAGC GTCTGCCCGG TGCGGTAAGG GATGTCGTAC
1701  ACGACCAGGG GGAGGCCGCC GTGCTCGGCC AGCGCGGCGA AATGAGCCAG
1751  GGTCCCCGCT TCCCCGGGGC GGATGTAGGG CGGCGCGGGG ACCAGCGCGG
1801  CGGCGACGTC ACCCCGGGCC GCCAGCTCTC GCAGGGCCGT GATGGCGGTG
1851  GCGGTGTCGT TGGTGCCCAC CCCGACGATG AGCGGTGCCC CGTGTGCCCG
1901  GCACGCGGCC GAGCAGACGC GGATCACCGT CTCTCTCTCC TCGGCGGTCA
1951  GTGTGGCGGC CTCGGCGGTC GTACCGAGGG CGACGAGCCC GGAGGCGCCG
2001  GCCGACAGCG CCTCGTCGGC GAGTCGGGCC AGCGCCTCGG GGGCCAGGCG
2051  CAGATCGTCG GTGAACGGAG TTACCAGGGG GACGTACAGG CCGTTGAAGA
2101  GCGGTTCGGT GGTCGGTTCG AGGCTCGATG CGAGGGTCAT GCTCTTACCC
2151  TGGCCCACGC CACTCGGTAG ATCCATTTCA GATTCCTGCC GTCACACCTA
2201  AGCTGAACTT ATGCTCGATG TCCGTCGCCT CCATCTGCTC CGCGAACTCG
2251  ACCGGCGGGG CACCATCGCC GCCGTGGCCG AAGCGCTGAC CTTCACCGCG
2301  TCCGCCGTCT CCCAGCAGCT CGGCGTGCTG GAGAGGGAGG CGGGCGTGCC
2351  GCTGTTGGAA CGCAGCGGCA GGCGCGTGGT CCTCACGCCC GCAGGACGCT
2401  CCCTCGTCGC ACACGCCGAC GCGGTGCTGA ACCGTCTCGA ACAGGCGGTC
2451  GCCGAGCTGG CGGGCGCACG GGACGGCATC GGCGGGCCGC TGCGCATCGG
2501  GACGTTCCCT TCCGGCGGCC ACACCATCGT CCCCGGCGCG CTGGCCGAAC
```

```
2551  TGGCCTCTCG TCACCCCGCG TTGGAGCCGA TGGTGCGGGA GATCGACTCC
2601  GCGCGCGTCT CCGACGGTCT GCGGGCCGGT GAGCTGGACG TGGCCCTCGT
2651  ACACGACTAC GACTTCGTAC CCGCGACGCC GGACACGACC GTGGACGAGG
2701  TGCCTCTGCT CGAAGAGCCG ATGTACCTCG TCACCCATGC CGCGGACACT
2751  GCCACGGACT CCGGCTCCGG GAGCACACTG GCAGCGCTGC TCGGGCCCTG
2801  TGCCGAGGTT CCGTGGATCA CGGCGCGGGA CGGCACGACC GGTCACGCGA
2851  TGGCTGTACG CGCCTGTCAG GCCGCCGGGT TCCAGCCCAG GATCCGCCAC
2901  CAGGTCAACG ACTTCCGCAC GGTGCTGGCT CTGGTCGCCG CCGGGCAGGG
2951  GGCCGGGTTC GTGCCGCGGA TGGCCGCCGA GCCGAGCCCC GCGGGCGTGG
3001  TGCTCACGAA GCTGCCGCTG TTCCGTCGCT CGAAGGTCGC GTTCCGTGCG
3051  GGCGGCGGTG CCCATCCGGC GATCGCCGCT TTCGTGGCCG CGGCGACGAC
3101  GGCGGTCGAA CGCATGGCGG GTTCACGAGG CCCGGCCGGC GGCTCTGAGT
3151  GAACCGGCCG ACCGTGGGAA TGTGTGTGCC CTGGGCCGCA CCATTCGTGG
3201  CCTGGTGACG TCCTGGCGAC GTCCTGACGT CCTGATGTCC GAACGAGAAG
3251  GCGATTTTCC GCGATGGCCG ATGACGCGTA CCTGTTCCTC CTCCCCGACC
3301  GGCACCCCCG ACTGGGAGCG GCCCTCGCCG CCGTCGGTGC CTTGGAATGC
3351  ACGGAAACCC CTGCGGTGCA CGCCTGGTTG CAGGCTCATG AGGCCTCCGT
3401  GTCCTCGGAA CAGGTCAGGA TTCTGCCCGC CGATGCCGAG ACACTCATCC
3451  CGAAGGACGC CGAGCGGCTG CCGGTGCCGT TGAGCGAGGA GGAGGCGCTC
3501  AAGGTCGAGC AGGAGTGCGC GCCCCAGACC GTCACGGACA TGGAGAGCGA
3551  ACTGCTCGCG TTCCGGGAGA CGACCCAGGA CTGGCAGGCC CTCGTGCACC
3601  GGGCCCTGAC CGCGGGCATC CCCGCGCAGC GCATCGCCCG GCTGACCGGA
3651  CTCGACCCGG AGGAGATCGG CCGCCTGTAG GCGCTAGCGG CCGCCCAGTG
3701  CGGACACCAG GATGGCGACC GTGACGGTGT TGAAGACGAA GGCGATGACC
3751  GTGTTCGCCG CCACGGTCCG TCGCATGTCG CGTGAGGTGA CGTCGACATC
3801  GGTGGTGCCG AACGTCGTCA TCGCGGCCAG GGCGAAATAG ACGTAGTCGG
```

3851 CCCAGGCGGG ACTCCGCTCC CCGGGGAATT CCAGTGCCCG CTCGTTCTCC

3901 ACGAGGTTGT CGGCCTGGAA GGTGACGGCG AAGGCCACGA CCACGCAGAT

3951 CCAGGCGGCG ACGACCAGGG CGAGCGCCAC CAGGGTGCGG GGGAGCGCGG

4001 AGAAGGTGGT GCTGAGGTGG CCGGGAAGCC ACAGCACCGC CACCACCAGC

4051 GCGGCAGCCG CGATGAAGAG CGAACCCCCG GGGCCGGGCG CGGTTCCGAG

4101 GACGTAACGC TGCAGGAATG TGCCGCGGGC TTCGCGCCGC GCCCAGGAGC

4151 GGACCTGCTC CGGAGCGACG CTCACGAAGA CGGTCATGGT GATGGCGAGG

4201 TAGGGCAGCA GGTAGGCGAA GAAGACGAGC ACGCCGACAT CCGCTGCCGA

4251 AATCCGCACC ACGGCGTCGA TGGGGAGGAC CACTGCCGCG CACGCCGCGA

4301 CGGCCAGGCT CACCGCCGAC CGGCGCCGTT CGGAAAGCCA GCGATGCACG

4351 GACGAGCCTC TCTGGTCGGG CGTCGGGCCT CGTGTGATCG TGACCGGCTC

4401 CGCGCCCGCC GAAAGCGCGG TGCGATCTCC TGCCCTCGAA CGAGCGAAAC

4451 GCTTGCGCCG GAAAGCCTCC CTGCTGATGC CGACGGCGGC GGCAGTGGCT

4501 GCGGATGCGG ATCGTGCGCT GTGCCCTGAC CCTGGATGGG GGGAGGAACG

4551 CAGAGAGGCA GGTGCGCCCA TGACGGTCAT GGACAAGCTC AAGCAGATGC

4601 TCAAGGGGCA CGAGGACAAG GCCGGCCAGG GAATCGACAA GGCGGGCGAC

4651 TTCGTCGACG GGAAGACGCA GGGCAAGTAC AGCGGTCAAG TCGACACGGC

4701 CCAGGACAAG CTCCGGGACC AGTTCGGCTC GGATCAGCAG GAGCCTCCGC

4751 AGAGGTAGGC AGCGTCAGGG CGGAATCGGT CCGGGCGACC GCTGACCGCT

4801 GATGCAGATG CCGCAGACGT CGGCCCCGCA CTCCTCCGGG TAAATCGGAG

4851 CGTAGGCGGG GCCGACGTGT GCGCGTGCGG CCTCGTCTCT GCCGCCCCTC

4901 TCCGCCCCGT CTCTGGCCCC TTGGTGCCAG TCTGACGGGA AAATGGCACC

4951 ACTTGGTGCC ACGCATGTGC CATGATGGCG TCATCGAGAG CGCGCTGCCC

5001 CGACTCGCGG GCAGGAAGGG CGCGTTCCGC GGAGTCGGCC GTCGGAGGGG

5051 TTGCATCATG GGGACAGCAC AGAGCCAGGA GCAGGCCGCC GCGCCCGGTG

5101 CCTGCGCCGC CTTCGTCCGC TTCGTGCTCT GCGGTGGCGG AGTGGGCCTC

```
5151  GCCTCCAGCT TCGCCGTGGT CGCCCTCGCC TCCTGGGTTC CCTGGGCGCT
5201  GGCCAACGCC CTGGTCGCCG TGGTCTCCAC CGTCGTCGCC ACCGAGCTCC
5251  ACGCCCGCTT CACCTTCGGT GCGGGCGGGC GCGCGACCTG GCGGCAGCAC
5301  GCGCAGTCGG CCGGGTCCGC GGCGGCCGCG TACGCGGTGA CCTGCGTGGC
5351  GATGTTCGTC CTGCAGCAGC TGGTGGCGGC GCCCGGCGCG GTGCTCGAGC
5401  AGGTCGTGTA CCTGTCGGCC TCCGCGCTCG CCGGTGTCGC GCGGTTCGTG
5451  GTGCTGCGCC TCGTCGTCTT CGCCCGGAAC CGCTCGCTGC CCGCCGCGGC
5501  CGCCGTGCGC ACCGCGCGTC CCGTGCGTCG CGTGCCGGCG CCCGTGCCCG
5551  CGACCGTGGC CCACGCCGCA TCGCGCCCGG CCGGCCCCGC GGCGCTCTGC
5601  CCCGCCGCAT GACTCCGTGC CCGCATGTTT GTGCCCCCGG TGCTCCGTGC
5651  GTCCGGGGGC GGGGTGGGCG TCGTGCCCGG GTGGTCCAGG GGTCACGCGG
5701  TGGTGTGTGC CAGTTCCTGG CCGAGGTGGT GGGCGAGCTG TGCGGGCGTG
5751  GGGTTCTCGA CGATGGCGAC CATCGCGATC TCCATCCCGG TCAGCGTCAT
5801  CAGTGTCTTG GTGAGCTCAA GGGCGGTGAG GGAGTTGAGA CCGTTCTCGA
5851  GGAAGTTGCT GTCGTCGCTG AGGGTGGTGT TCAGAAGGGT GCCGGCCTGG
5901  GTGCGGATGG TGTCGGTGAG GAGCTTCTCG CGCTCCTCGG GGGTGGCCGC
5951  GGCGAGCTGC TTCTCCAGCT CGGTGGCGTC CTGGCCGGAG GTGTGGTCGG
6001  TGCTGGTCAT GACTGCTCCT GTGTGAGTGA GGTGTTGGCG GGGGTCACAC
6051  CGCGGCGTGC GCGGTGTGGT CGTGCAGCCA GTAACGCGTG GCCTGGAAGG
6101  AGTACGTCGG GAGGTCGATG GTCCGGGGGT GGGGGGTGCG CCGGACGAGA
6151  GGGGTCCAGT CGACGGTGCC GCCCGTGGTG TGAAGCCGCG CGAGGGCGGT
6201  CAACAGGGCG CTTACGGCGG AGGTTTGCGT GCCTTCCGGT GAGAGCGCGC
6251  CCAGGTGGAG AAGCGTGTGG GTCTCGGGGG TGGGGGGTGC GGTGGGGGCC
6301  GGCGAGGTGA GGTGGTGGTG CCAGTAGTCG GCGGAGGCGA TGGGGGTGTC
6351  GGCCGGGGCA GTGCTGGTGA GCGTGAGCGT GGCGCGTTGG AACGTCAGCT
6401  GCTTCAGCAC GGGCTCGTAG GCGTCGGGCG GAGCCGGTTG TTCACCCTCG
```

6451 GCGGCCTGGG CGGCAGCGGC GTGGGCGGCG GCCAGGCGGC ACGCGTCGTC

6501 GAGGGTCAGG ATTCCCGCGG CGTACGCGGC GGCGATGTGG CCGACACCGT

6551 CGCCGGTGAG GGTGTGGGGG CGTACCCCCG TTTCCAGGAG CAGCCGCGCG

6601 AGCGCGGTGT GGACCGCGAA GCGCGCCAGT TCGGAGTGGG GAGTGGGGAG

6651 GGGAGTCGGC AGATGGGTGT CGAGGAGCGC GCGCGCTTCG TCGAAGGCGG

6701 ACGCGAAGAG CGGGAACGCC GAGTGGAACT CGGCACCTCC GAAAGCCGCG

6751 CCGAATGTAG CGCCGAATGT CGCGCCGGGT TTGGCTCCGG GTGCCGCCCC

6801 CGTCGTCACC CCGTCGGCCG GGCGGCCGTC GAAGTGCCAG GCGATCTTCT

6851 TCGGGCCGGC CCCGGGCGTG GACCTGACCA GGTCCGGGTG GTCCTCTCCG

6901 GCGGCCAGGG CGCGGGCGGC GGCGAGGAGT TCGGTGTGGT CGGTGCCGGT

6951 GAGGACGGCG CGGTGTTCCA GGGGGCTGCG GGTGGCGGCG AGCGAGTAGG

7001 CGACCTCGGC GGGGGAGGGC GCGGGGTCGG TGGCCGCCAG GTGGGTGACG

7051 AGGGCCTTCG CCTGTGCCCG CAGGGCCTCG GGTGTACGAG CGGACAGGCT

7101 CCAGGCCACC GGGAGTTCCG GGGCAACGGG CGACGTCTGG TCGCGGGCGG

7151 CATCCGGCAC CGGAGCCTCG TCCACCGGCG GCTCTTCGAG GATGAGGTGC

7201 GCGTTCGTGC CGGACGTGGC GAAGGCGGAG ATGCCGACCC GGCGGGGCTC

7251 CTCGCGGCGG GGCCAGTCGA CCGCCTCGGT GAGCAGCCGT ACCGCGCCCT

7301 TCTTCCAGGC GGCGAGGGGC GTCGGGCGGT CGACGTGGAG GGTCGGCGGC

7351 AGGGTGCCGT GCCGGAACGC CTGGACCATC TTGATGAGCG CGGCCGCACC

7401 CGCGGCCCCC TGCGTGTGCC CCGTGTTGGA CTTGACGGAG CCGAGCCACA

7451 GGGGCCGGTC GGGGGAGCGG TCGGCGCCGT AGGTGGCGAG GAGGGCCTGG

7501 ACCTCGATGG CGTCGCCGAT GGGGGTGCCC GTCCCGTGCG CCTCGACGGC

7551 GTCGATCTGG TCCGGGGTGA GCCCGGCGTC GGCGAGGGCG GCGCGGATCA

7601 CATGCTGCTG GGAGGGGCCG TTGGGGGCGG CGAGGCCGTA TCCGGCGCCG

7651 TCCTGGTTGA CCGCGGAGCC GCGGATGACG GCGAGCACCG GGTGGCCGTT

7701 CTTCCTGGCG TCGCCGAGCC GCTCAAGCAG GACGAGGCCG ACGCCTTCAC

7751 CGAGGCCCAT GCCGTCGGCC GCGGCGGCGA ACGGTTTGCA ACGGCCGTCC

7801 TGCGCGAGCG ACTTCTGGTG GGCGAAGGCG TGGAAGGTGT GCGGCGTCGA

7851 CATGACGGTG CCGCCGCCGG CGAGGGCGAG GCCGCACTCC CCGCGGCGCA

7901 GCGCCTGGCA GGCCAGGTGG AGGGCGACCA GGGAGGACGA GCAGGCCGTG

7951 TCCACGCTGA TGGCGGGGCC CTCGAGGCCG AGGGCGTAGG CGATGCGGCC

8001 GGAGACGAGG CTGCCGGACG TGCCGCCGCC CAGATAGGGC AGCAGCTCGT

8051 CGGGCGCGGT CTCGAGCCGT GTCGCGTAGT CGTGCCCGGT GGCGCCGACG

8101 TAGACGCCGG TGAGGGTGGA GCGCAGGGTG TGCGGGGCGA TGTGGCCGCG

8151 TTCGACGGTC TCCCACGCGA GGTGGAGCAT GAGGCGCTGG AGGGGTTCGG

8201 TGGCCACGGC CTCGGTGTCG CTGATGTCGA AGAAGCCCGC GTCGAAGCCG

8251 GCCGCGTCGT CCAGGAACCC GCCGAGCTCC GCGTACGGGC GTTCCTCGGG

8301 GAGTTCCCAG GCGCGGTCGT CGGGGAAGCC GGTGACGGCG TCGCGGCCCT

8351 CGGACACCAG ATCCCACAGG TCGTCCGGGG TGCGGGTCTT GCCGGGCAGC

8401 CGGCAGGCCA TGGAGACGAC GGCGATCGGC TCGTGCTGTG CGGCCTTCAG

8451 TTCGCGCAGC TGCTGCTGGG CCTGGTGGAG CTCGGCCGTC GTCCACTTGA

8501 GGTATTCGAC GAGCTTCTCT TCGTTCGCCA CGGGAATGGT CAGCCTTCCT

8551 GTTCTCGCGC GTGAAGCCTC AGGTGGGACG AGGTCGGGCA AGGTGGGCAG

8601 GCAGGAGCCG CGCGCTGTGG GTGCCAGGGT CGCCGCGGCT GCTTAAGCGG

8651 GTCTAACTCC CGCCTTGCCG CCGGGCATCG CCTCGCACGA GCGGGCCAGC

8701 AGCAGGAGGT CGGCGGCGAT CTCGTCGGGT GCGCCGGCGT GCAGATCGTG

8751 GTCGGAGCCC GGGTACCAGC GCACGCTCAC CTGCTCCAGG GCCGCCTCGG

8801 CGGCGGCCAC CCAGGCCCGT ACCTGGTCGG ACAGTTGGGG GATGGCGGGG

8851 ATGAGGGGCA GCAGCCGCAC CGGCACGGTG ACCTTGGGAT ACCAGTCGGC

8901 CGGTGCCTCC CGTTGCAGGC CGGCGACGAT CGACATGACC TGTGTCGAGG

8951 TCAGGCGGGG GATGAGCAGG CCGTCCGGCC CGACGCGGTA GTCCGCCAGG

9001 CGTGCCTCGA TGGACGTGGG CGACCAGTCG GGATGGGTGG CCCGCAGGTA

```
9051  GGCGCGCATG TCGGCGGCGC TGGTGGTGCC CTGCTGGGCG CGCCGGACCA
9101  CGTCGGCGGT GCGCTCCCAG AAGGCGCGCA TCACCGGTCC GTCGAACTCG
9151  TACCAGCCGC CGTCGATCAG GGCGAGACCG GCCACCAGGT CCGGGTGCTC
9201  GGCCGCCAGG CGCAGCGCGA GGTGCGCGCC CCAGGAGTGC CCGGCCACCA
9251  GTGCGCCGGA CAGGTCGAGG GCGGTGACGG CCGCCACCAG GTCGGTGACG
9301  ACCGTCGCGT TGTCGTACCC GTCGGGCGGG GTGTCCGACT CGCCGTGGCC
9351  GCGGTGGTCG ACGGCGTAGG CCGGGTGTCC GGCGGCGGCG AGACGGGCGG
9401  CGACCTCGTC CCACATCCGG GCGTTCGACA GCATGCCGTG CAGCAGCAGG
9451  AACGGACGGC CCGGGGCTCC CGGCCCGTCC GCGGGCCGGT ACCTGACATT
9501  GAGGGAGACG GTCTGCGACA CGGGGATGCG GAGGTTCTTC ACAGGCGGGC
9551  CCTTGTGATC CCTTGTGCTG GGGGAGGAAA GCGGGGGCGG CACGCTCAGG
9601  GGCGCTGCGC GGTCGCGAAG ATGTATCCGA GCTCGGGCAT CTTGCCGAGG
9651  GCCGCCTGGT TGTGCAGGAA CAGCTCGTAT CCCTCTACGC CGATGATGTC
9701  GACGTACTCG TCCCGGTGGG CGCGGATCCA CTCGACGTAA CCGTCGTAGG
9751  TCTTGGCGGT CTCGCGGGTG ATGTCGGTCA GTTCGAGGAC GGTCCAGCCG
9801  GCGGCGCGGA AGATGTCGGG GTAGTCCCCG ATGTCGGTGA GCGCGGCGTA
9851  GATCGTGGTG TCGCTGACGG TGGCGGTCCG GGGCCGGCTG GGATCGGGGT
9901  TGAGGTAGAC CATGTCGGCG ATCGGCATCC GCGCGCCGGG CTTCACGACG
9951  CGGTGGGCCT CGGTGAGCAC CTGCTGCTTG TCCGGCATGT GCAGCATGGA
10001 CTCCAGGGCC CAGCAGTGGT CGAAGGAGCC GTCGTCGAAC GGCAGGTTCA
10051 TGGCGTCGAC CTGCTCGAAG CGGACCCGGT CGGCGAGGCC GGCCTCGCGC
10101 GCCCGGCGGT TGCCGCGCTC GACCTGGCGG GCGCTGACGG AGATGCCGAC
10151 CACCTCGACG TCGCGGGCGC GGGGCAGCTG CATGGCCGGG GTGCCGTTGC
10201 CGCAGCCGAT GTCGAGGACG CGGTCGCCGG GGGCCGGGTC GAGGCGGCGG
10251 ATCATCTCGT CGGTCATCTG GACCATGGCC TCGTCGAACG TGGCCTGCTG
10301 CTCGCCGCCG TCGAACCAGT AGCCGTAGTG CAGATTGCCG TCTCCGAGCT
```

```
10351  GAGTCATCAG GTCGAAGACC TTGTGGTCGT AGTAGTGGCC GATGTCGCTG
10401  GGCTCGGGGG CGACGGTCTT GTTCACCGTC GGGGGCTTCT TGGTCGTCGC
10451  GTTCTTCGTC ACGGCTTCAG CGTCACCGTG CGGCGGCAGG CGCCACAACC
10501  CCACCCCCGC CCCTCAAAAG CCCCTATGGG CCCTCCTCGA CCGCCCCTAG
10551  GGAGCTGCTC TTGACGCGTT CCATACGGAA CGGGTGGTAC CCCTCCGAAA
10601  AAAATGAGAG TACGCTCCCA CTAGATATTG AGCTCTCTTT AGGAGGTCGA
10651  CTCCCATGTC TGCTGATCTG GGTGCGCGGC GGTGGTGGGC CGTCGGTGCT
10701  CTCGTACTCG CCTCGATGGT CGTGGGCTTC GATGTGACGA TCCTGAGCCT
10751  GGCGTTGCCC GCCATGGCCG ACGACCTCGG CGCGAACAAC GTCGAGCTGC
10801  AGTGGTTCGT GACGTCGTAC ACGCTGGTGT TCGCGGCCGG CATGATCCCG
10851  GCCGGCATGC TCGGTGACCG GTTCGGACGC AAGAAGGTCC TGCTCACCGC
10901  CCTGGTGATC TTCGGTATCG CCTCGCTGGC CTGTGCCTAC GCGACGTCCT
10951  CCGGCACCTT CATCGGCGCG CGTGCGGTGC TCGGTCTGGG CGCCGCGCTG
11001  ATCATGCCGA CGACGCTGTC GCTGCTGCCG GTCATGTTCT CCGACGAGGA
11051  GCGGCCGAAG GCCATCGGAG CGGTGGCCGG TGCGGCGATG CTCGCCTATC
11101  CGCTCGGCCC GATCCTCGGC GGCTACCTGC TCAACCACTT CTGGTGGGGC
11151  TCCGTCTTCC TGATCAACGT GCCGGTGGTG ATCCTCGCCT TCCTCGCGGT
11201  CTCCGCCTGG CTGCCCGAGT CCAAGGCCAA GGAGGCCAAG CCGTTCGACA
11251  TCGGCGGCCT GGTGTTCTCC AGCGTCGGTC TCGCCGCGCT GACCTACGGC
11301  GTGATCCAGG GCGGCGAGAA GGGCTGGACG GACGTCACCA CGCTGGTGCC
11351  GTGCATCGGC GGTCTGCTCG CCCTCGTGCT GTTCGTGATG TGGGAGAAGC
11401  GGGTGGCGGA CCCGCTGGTC GACCTCTCGC TGTTCCGCTC GGCCCGGTTC
11451  ACCTCCGGCA CCATGCTCGG CACCGTCATC AACTTCACGA TGTTCGGCGT
11501  GCTCTTCACG ATGCCGCAGT ACTACCAGGC GGTCCTCGGC ACCGACGCGA
11551  TGGGCAGCGG CTTCCGGCTG CTCCCGATGG TCGGCGGTCT GCTCGTGGGT
11601  GTGACGGTCG CCAACAAGGT CGCCAAGGCC CTCGGCCCGA AGACCGCGGT
```

11651 CGGCATCGGC TTCGCCCTCC TCGCCGCCGC CCTGTTCTAC GGCGCCACCA

11701 CGGACGTCAG CAGCGGCACC GGCCTGGCGG CCGCCTGGAC CGCGGCCTAC

11751 GGACTCGGCC TCGGCATCGC CCTGCCGACC GCCATGGACG CCGCCCTCGG

11801 CGCGCTCTCC GAGGACTCCG CCGGCGTCGG ATCCGGCGTC AACCAGTCCA

11851 TCCGTACCCT CGGCGGCAGC TTCGGCGCGG CCATCCTCGG TTCCATCCTC

11901 AACTCCGGCT ACCGCGGCAA GCTCGACCTC GACGGCGTGC CCGAGCAGGC

11951 ACACGGCGCG GTCAAGGACT CCGTCTTCGG CGGCCTCGCG GTGGCCCGGG

12001 CGATCAAGTC CAACGGACTG GCCGACTCGG TGCGTTCCGC GTACGTCCAC

12051 GCCCTGGACG TGGTGCTCGT GGTCTCCGGC GGCCTCGGAC TGCTGGGTGT

12101 GGTGCTGGCG GTGGTGTGGC TGCCCCGCCA TGTTGGTCAG AGCACCGCCA

12151 AGACAGCAGA ATCTGAGCAT GAAGCCGCAG ACGCAGTCTG ACCAGGGCAA

12201 AACAGTGCCT GGTCTGAGAG AACGCAAGAA GGCCCGGACG AAGGCCGCGA

12251 TTCAGCGGGA GGCGGTGCGC TTGTTCAGGG AACAGGGCTA CACCGCCACG

12301 ACCATCGAGC AGATCGCCGA AGCCGCCGAG GTCGCTCCCA GCACCGTCTT

12351 CCGCTACTTC GCGACCAAGC AGGACCTGGT CTTCTCGCAC GACTACGATC

12401 TGCCCTTCGC GATGATGGTC CAGGCCCAGT CACCCGACCT GACGCCGATC

12451 CAGGCCGAGC GGCAGGCCAT CCGCTCGATG TTGCAGGACA TCAGCGAGCA

12501 GGAACTGGCC CTGCAGCGCG AGCGGTTCGT CCTGATTCTC TCCGAGCCGG

12551 AGCTCTGGGG CGCCAGCCTC GGCAACATCG GCCAGACCAT GCAGATCATG

12601 AGTGAGCAGG TGGCCAAACG GGCCGGGCGC GACCCGCGGG ACCCCGCGGT

12651 CCGCGCCTAC ACCGGAGCCG TGTTCGGAGT GATGCTCCAG GTCTCGATGG

12701 ACTGGGCCAA CGATCCGGAC ATGGACTTCG CGACCACGCT GGACGAGGCA

12751 CTCCACTACC TGGAAGACCT GCGGCCCTGA CCGAAGGGGC GGGCGCACAC

12801 CACAGAGCCC GCCCCGGCCA GACGTCGTAC GAGGCGCCAT CGGCCGTCGC

12851 GTACGACCCC CGCGCCCCGG ATTCCCCCGC GGGGCGCGGG GTCAAGGGAA

12901 AAGAGACGAC CGCACGCGGC CACTGTTCCC CCGGCTGCCG CGTCCGGTCC

12951 AACCTGGCGT GCTCCGGCTT CCCTCGACGG AGCACGCCAG GGGTCTGTCC

13001 GGCCCTCTCC CGGCGGCTCC CGTCAGACGC CCGGCCCCGC CGTCAGCGCC

13051 TCGGTCACGA CGGCCGCCAC CTGCTCCTGA CAGCCGTCGA GGTAGAAGTG

13101 CCCGCCGGGC AGCACCCGCA GATCGAACGC CGCGCCGGTC CGCTCCCGCC

13151 ACGTGGCGGC CTGCTCCGGC GACGTCCGCT CGTCGGCGTC CCCGATCAGC

13201 GCCGTGATCG GGCAGTCGAG CCGGCCGGGT CCCGGCGCCT CGTAGGTGGC

13251 CACGGCCCGG TAGTCGGCCC GCAGCGCGGG CAGGACGAGC TCCTGCAGCT

13301 CGGGACTGCG GAAGAACCGC TCGTCGGTGC CGCCCATCGC CCGCAGATGG

13351 GCCAGGATGT CCGCGTCCCC GAACGCCCCC GAACGCCCCG CGGGACGGTA

13401 GGGCCGCGCG AGCCCCCCGG AAACGAACAG GTGCACGGGA AGGCCGGGCC

13451 CGGCCGGTCC CCGCAGCCGC CGCGCCACCT CGAACGCCAC GATGGCGCCC

13501 AGGCTGTGCC CGAACAGCGC GAATGGCTTC CCGTCGCACG GCAGGTGGGG

13551 CACGACGCCG TCGGCGAGCT CGGCCACCGA CGCCAGGCAC GGCTCCGCAT

13601 GACGGTCCTG CCGCCCCGGA TACTGCACGG CGAGCACCTC GACGCCGGGC

13651 GCGAGCAGCC CGGAGAGCCC GAAGTAGTAA CTCGCCGAAC CGCCCGCGAA

13701 CGGAAAGCAG ACCAGCCGCA CCGGCGCCTC TGCCGCAGCG TGGTACCGCC

13751 GCAACCACAC CCCGTTTCCG GTGGCTGCAC CGAACTCGTC ACCGATCTGT

13801 GGTGCCCGCG CCGCCGTGCC CCTGTCCATC GTTCTCCCTC TCCTCGCGTC

13851 GCTCCGCGGG CGCTGTCCTG CCCCGCCCCG AAAGCCCGAT GCCGGCCAAG

13901 CCCCGATGCT GGCCAAACCC CGATGCCGGC CAAGCCCCGA TGCTGGCCGC

13951 GGCCCATAGC GCCCGGCTAA AGCCGCAGGC GGCTAGCCGG GGTTTGGTTC

14001 GCCTTTAGAC AGCCCACCCA CGATGAGCCC GGTACTCGAA GCGATCTCCG

14051 ATTTCGGACC GGGAGCGCCG TTGATGTTTT GTGGCAGCCA GTTGTTCAGC

14101 GCCCGACCGC AGCTGACGTG ATGGCCGCAT CCGCGTCAGC GTCCCCCTCG

14151 GGACCGAGCG CAGGACCCGA CCCGATCGCC GTCGTCGGGA TGGCCTGCCG

14201 CCTGCCCGGA GCACCTGACC CCGACGCGTT CTGGCGGCTG CTCAGCGAGG

14251 GGCGCAGCGC GGTGAGCACC GCACCGCCCG AGCGGCGGCG AGCCGACTCC

14301 GGCCTCCACG GGCCGGGCGG CTACCTGGAC CGGATCGACG GCTTCGACGC

14351 GGACTTCTTC CACATCAGCC CGCGCGAGGC CGTGGCGATG GACCCCCAGC

14401 AGCGGCTGCT CCTCGAACTG AGCTGGGAGG CCCTCGAAGA CGCGGGCATC

14451 CGGCCGCCCA CCCTGGCGCG CAGCCGCACC GGCGTCTTCG TCGGCGCGTT

14501 CTGGGACGAC TACACCGACG TCCTGAACCT GCGGGCGCCG GGCGCCGTCA

14551 CCCGCCACAC CATGACCGGC GTGCACCGCA GCATTCTGGC CAACCGCATC

14601 TCGTACGCGT ACCACCTGGC CGGTCCGAGC CTCACCGTCG ACACCGCACA

14651 GTCCTCCTCG CTCGTCGCCG TCCACCTGGC CTGCGAGAGC ATCCGCAGCG

14701 GCGACTCCGA CATCGCCTTC GCGGGCGGCG TCAACCTCAT CTGCTCGCCG

14751 CGCACCACCG AGCTGGCCGC GGCCCGCTTC GGCGGTCTCT CGGCCGCAGG

14801 CCGCTGCCAC ACCTTCGACG CCCGCGCCGA CGGTTTCGTA CGCGGCGAGG

14851 GCGGCGGCCT CGTGGTGCTC AAGCCCCTCG CGGCGGCACG GCGCGACGGC

14901 GACACGGTGT ACTGCGTGAT CCGGGGGAGC GCCGTCAACA GCGACGGTAC

14951 GACCGACGGA ATCACCCTGC CCAGCGGGCA GGCGCAGCAG GACGTGGTGC

15001 GCCTCGCCTG CCGACGGGCG CGGATCACGC CGGACCAGGT GCAGTACGTC

15051 GAACTGCACG GCACCGGCAC GCCCGTCGGG GACCCGATCG AGGCCGCCGC

15101 GCTCGGCGCC GCCCTCGGGC AGGACGCCGC CCGCGCCGTG CCGCTGGCCG

15151 TCGGCTCCGC CAAGACGAAC GTCGGCCACC TCGAAGCCGC CGCCGGAATC

15201 GTCGGACTGC TCAAGACCGC CCTGAGCATC CACCACCGGC GGCTGGCGCC

15251 GAGCCTGAAC TTCACCACCC CCAATCCGGC CATCCCGCTC GCCGACCTCG

15301 GCCTGACCGT CCAGCAGGAC CTGGCCGACT GGCCGCGCCC CGAACAGCCC

15351 CTGATCGCCG GGGTGTCGTC CTTCGGCATG GGCGGCACGA ACGGTCACGT

15401 TGTCGTGGCG GCGGCGCCCG ATTCGGTGGC GGTACCTGAG CCGGTGGGGG

15451 TGCCTGAGCG GGTGGAAGTG CCTGAGCCGG TGGTGGTTTC TGAGCCGGTG

15501 GTGGTGCCGA CGCCATGGCC CGTGAGCGCT CACAGCGCTT CCGCGCTGCG

```
15551  CGCGCAGGCC GGTCGCCTGC GGACGCACCT CGCCGCCCAC CGCCCCACCC
15601  CCGACGCCGC GCGGGTCGGC CACGCGCTCG CCACCACCCG TGCGCCCCTC
15651  GCCCACCGCG CGGTCCTGCT CGGCGGCGAC ACCGCCGAAC TGCTGGGCTC
15701  CCTGGACGCG CTGGCCGAGG GCGCGGAGAC CGCGTCCATC GTGCGCGGCG
15751  AGGCGTACAC CGAGGGCAGG ACGGCCTTCC TCTTCAGTGG GCAGGGAGCG
15801  CAACGCCTCG GCATGGGGCG GGAGTTGTAT GCCGTGTTCC CCGTCTTCGC
15851  CGACGCTCTC GACGAGGCGT TCGCCGCCCT GGACGTACAT CTGGACCGCC
15901  CACTGCGCGA GATCGTCTTG GGCGAGACCG ACTCGGGTGG GAACGTCTCG
15951  GGTGAGAATG TCATCGGCGA GGGTGCCGAC CATCAGGCAC TCCTCGACCA
16001  GACCGCCTAC ACCCAGCCCG CGCTCTTCGC GATCGAGACG AGCCTGTACC
16051  GGCTGGCAGC CTCCTTCGGC CTGAAGCCGG ACTACGTCCT CGGCCACTCG
16101  GTCGGCGAGA TCGCCGCCGC GCACGTCGCC GGTGTCCTCT CGTTGCCGGA
16151  CGCGAGCGCT CTGGTGGCCA CGCGGGGACG GCTCATGCAG GCGGTTCGCG
16201  CGCCCGGCGC GATGGCCGCG TGGCAGGCCA CGGCGGACGA GGCGGCCGAA
16251  CAGCTCGCCG GGCACGAGCG GCACGTCACC GTGGCCGCCG TCAACGGCCC
16301  CGACTCCGTG GTCGTCTCCG GCGACCGCGC CACCGTCGAC GAACTGACCG
16351  CCGCCTGGCG GGGACGCGGC CGCAAGGCCC ACCACCTGAA GGTCAGCCAC
16401  GCCTTCCACT CCCCGCACAT GGACCCCATC CTCGACGAGC TGCGCGCGGT
16451  CGCCGCCGGC CTGACCTTCC ACGAGCCGGT CATTCCCGTC GTCTCCAACG
16501  TCACCGGTGA ACTGGTGACC GCGACCGCGA CCGGGAGCGG CGCCGGGCAG
16551  GCCGACCCCG AGTACTGGGC GCGGCATGCG CGCGAGCCCG TGCGGTTCCT
16601  GTCCGGGGTG CGGGGGCTGT GCGAGCGCGG GGTGACCACG TTCGTCGAGC
16651  TCGGCCCGGA CGCACCGCTG TCCGCGATGG CCCGCGACTG CTTCCCCGCC
16701  CCCGCGGACC GGAGCCGTCC GCGCCCCGCC GCCATCGCCA CATGCCGCCG
16751  CGGGCGCGAC GAGGTGGCCA CGTTCCTGAG GTCGCTGGCC CAGGCGTACG
16801  TCCGCGGCGC CGATGTCGAC TTCACCCGGG CCTACGGCGC CACCGCCACG
```

```
16851  CGCCGCTTCC CCCTCCCCAC GTATCCCTTC CAGCGCGAGC GCCATTGGCC
16901  TGCCGCTGCC GGGGTGGGGC AGCAGCCGGA GACCCCGGAA CTTCCGGAAT
16951  CCTCGGAGTC CTCGGAGCAG GCAGGGCATG AGCGGGAGGA GGGGGCGCGC
17001  GCGTGGGGCG GGCCTGAAGG GCGGCTTGCC GGGCTCTCCG TGAACGACCA
17051  GGAGCGGGTC CTCCTCGGCC TGGTCACCAA GCACGTGGCC GTCGTGCTCG
17101  GGGACGCCTC GGGCACGGTA CAAGCCGCCC GCACCTTCAA GCAGTTGGGC
17151  TTCGACTCGA TGGCCGCCGC CGAGCTGAGC GAACGGCTCG GCACGGAGAC
17201  GGGCCTGCCG TTGCCCGCCA CCCTCACCTT CGACTACCCG ACCCCTCTGG
17251  CCGTCGCCGC GCACCTGCGC GCGGAGCTCA CCGGTACGCC CGCCCCGGCC
17301  GGCTCCGCGC CCGCCACGGG CGCCCTCGGC GCGGGTGACC TCGGCACGGA
17351  CGAGGACCCG GTCGCCATCG TGGCCATGAG CTGCCGCTAT CCCGGCGGCG
17401  CAGGCACGCC CGAGGACCTG TGGCGGCTGG TCGCGGACGG CGCCGACGCG
17451  ATCGGAGACT TCCCCACCGA CCGCGGCTGG GACCTGGCGC GGCTGTTCCA
17501  CCCCGACCCC GACCGGTCGG GCACCAGCTG CACGCGGCAG GGCGGATTCC
17551  TGTACGACGC CGCCGACTTC GACGCCGAGT TCTTCGACAT CAGCCCGCGC
17601  GAGGCCCTGG CCGTCGACCC GCAGCAGCGG CTGCTCCTCG AGTGCGCCTG
17651  GGAGGCCTTC GAACGGGCGG GCCTGGACCC GCGGGCGCTC AAGGGCAGCC
17701  CCACCGGCGT GTTCGTCGGC ATGACGGGGC AGGACTACGG CCCCCGTCTG
17751  CACGAGCCGT CCCAGGCCAC CGACGGCTAT CTGCTGACCG GCAGCACGCC
17801  GAGCGTGGCC TCGGGCCGCC TGTCGTTCAG CTTCGGCCTT GAGGGGCCCG
17851  CCCTGACGGT GGACACGGCC TGCTCGTCGT CGCTGGTCAC GCTCCATCTC
17901  GCGGCGCAGG CGCTGCGGCG CGGCGAGTGC GACCTGGCCC TCGCCGGCGG
17951  CGCCACCGTC CTGGCCACGC CGGGCATGTT CACCGAGTTC TCGCGGCAGC
18001  GGGGCCTGGC CCCCGACGGC CGCTGCAAGC CGTTCGCGGC GGGCGCCGAC
18051  GGCACGGGCT GGGCCGAGGG CGTGGGCCTG GTCCTCCTCG AAAGGCTCTC
18101  CGAGGCCCGG CGCAAGGGGC ACGCCGTCCT CGCGGTGATC CGGGGTTCGG
```

```
18151  CGATCAACCA GGACGGCGCG AGCAACGGCC TGACCGCGCC CAACGGCCCC
18201  TCGCAGCAAC GCGTCATCCG TGCCGCGCTC GCGGCCGCCC GGCTCACCGC
18251  GGACGAGGTC GACGTAGTGG AGGCGCACGG CACCGGCACC ACGCTCGGCG
18301  ACCCGATCGA GGCGCAGGCC CTGCTCGCCA CGTACGGCCA AGGGCGTTCG
18351  GCGGAGCGGC CGTTGTGGCT CGGGTCGGTG AAGTCGAACA TCGGTCACAC
18401  GCAGGCCGCC GCGGGTGTCG CGGGCGTCAT CAAGATGGTG ATGGCGATGC
18451  GCCACGACCT GCTCCCCGCC ACCCTGCACG TCGACGAGCC GAGTGGCCAC
18501  GTGGACTGGT CCACCGGCGC GGTGCGACTG CTCACCGAGC CGGTCGTCTG
18551  GCCGCGCGGC GAACGTCCGC GCCGCGCCGC GGTGTCGTCC TTCGGCATCT
18601  CCGGCACGAA CGCGCACCTG GTGCTCGAAG AGGCGGGGCA GGACGAGTAC
18651  GTTGCGGGAG CCGCCGACGA CGCCGGGCCG GTGGACGGTG CTGTGCTGCC
18701  GTGGGTGGTT TCCGGACGGA CCGGAGCGGC GCTGCGCGAA CAGGCCCGCC
18751  GTTTGCGTGA GTTGGTGACC GGCGGCTCGG CCGATGTCTC TGTGTCCGGG
18801  GTGGGCCGGT CGCTGGTCAC CACGCGGGCG GTGTTCGAGC ACCGGGCCGT
18851  GGTCGTGGGC CGCGACCGGG ACACGCTGAT CGGCGGCCTC GAGGCCCTTG
18901  CGGCGGGTGA CGCGTCGCCG GACGTCGTGT GCGGGGTCGC GGGCGATGTC
18951  GGCCCCGGCC CGGTGCTGGT GTTCCCCGGG CAGGGCTCGC AGTGGGTGGG
19001  CATGGGAGCC CAACTCCTTG GCGAGTCCGC GGTGTTCGCG GCGCGGATCG
19051  ACGCGTGCGA GCAGGCGCTG TCCCCGTACG TCGACTGGTC ACTGACAGAG
19101  GTCCTGCGCG GGGACGGGCG CGAACTGTCG CGCGTCGACG TCGTCCAGCC
19151  CGTGCTGTGG GCGGTGATGG TCTCGCTCGC CGCCGTCTGG GCGGACCACG
19201  GCGTCACCCC GGCCGCCGTC GTCGGGCACT CCCAGGGAGA GATCGCCGCT
19251  GTGGTCGTCG CCGGCGCGCT CACCCTGGAG GACGGCGCCA AGATCGTGGC
19301  CCTGCGCAGC CGGGCGCTGC GTCAGCTCTC GGGCGGGGGC GCCATGGCCT
19351  CCCTCGGGGT GGGCCAGGAA CAGGCAGCCG AACTCGTCGA GGGCCACCCC
19401  GGAGTGGGCA TCGCCGCCGT CAACGGCCCG TCATCGACCG TCATTTCAGG
```

19451 CCCGCCCGAG CAAGTCGCCG CCGTCGTCGC CGACGCCGAG GCGCGCGAGC

19501 TGAGAGGCCG CGTCATTGAC GTGGACTACG CCTCGCACAG CCCCCAGGTC

19551 GACGCCATCA CCGACGAACT CACCCACACC CTGTCCGGCG TCCGCCCCAC

19601 CACGGCCCCG GTGGCGTTCT ACTCGGCCGT GACCGGAACC CGCATCGACA

19651 CGGCGGGCCT CGACACCGAC TACTGGGTCA CCAACCTGCG CCGCCCGGTC

19701 CGGTTCGCCG ACGCCGTCAC CGCGCTCCTC GCCGACGGCC ACCGGGTCTT

19751 CATCGAGGCC AGCAGCCACC CCGTCCTCAC CCTCGGCCTC CAGGAGACCT

19801 TCGAGGAGGC CGGGGTCGAC GCCGTCACCG TCCCCACCCT GCGGCGCGAG

19851 GACGGCGGCC GGGCACGCCT GGCCCGCTCG CTGGCACAGG CCTTCGGCGC

19901 CGGGTGCGCG GTGAGGTGGG AGAACTGGTT TCCGGCCACC GGTACGTCCA

19951 CCGTGGAGCT GCCGACGTAC GCCTTCCAGC GTCGCCGTTA CTGGCTGGAG

20001 GCCCCCACGG GCACCCAGGA CGCGGCGGGC CTGGGCCTCG CCGCTGCGGG

20051 GCACCCGCTC CTCGGGGCGG CCACCGAGAT CGCGGACGGC GACATCCGCC

20101 TGCTCACCGG CCGTATCAGC AGGCACAGCC ACCCCTGGCT CGCTCAGCAC

20151 ACCCTCTTCG GTGCCGCGGT CGTGCCCGCC TCCGTCCTCG CGGAATGGGC

20201 GCTGCGCGCC GCCGACGAGG CCGGCTGCCC GCGTGTCGAC GACCTCACGC

20251 TGCGCACCCC GCTGGTGCTG CCCGAGACCG CGGGCGTGCA GGTGCAGATC

20301 GTGGTCGGCC CGGCCGACGC GCGGGACGGG CACCGCGACT TCCACGTCTA

20351 CGCCCGCCCC GACGGCAAGG ACGCCTCTGA GGGCGAGGGC ATCGCCGAGG

20401 GCGAGGGTGC CTCTGAGGGC GAGGGTGCCT CCGGCGGCAC CGATGCGCCG

20451 TGGACCTGCC ATGCCGACGG CCGACTGGTC GCCGAGCCCA CCGGCACGGC

20501 CTCGGAGGAC TCCCCGGACA CGGTGTGGCC GCCGCCCGGC GCCGAACCCG

20551 TCGACCTGGG CGACTTCTAC GAGCGGGCCG CCGCCACCGG AGTCGGCTAT

20601 GGACCGGTCT TCACGGGGCT GCGCGCCCTG TGGCGGCGGG ACGGCGAGCT

20651 GTTCGCCGAG GCGGTGCTGC CGCAAGAAGC CCCGGAAACC GCCGGGTTCG

20701 GCATGCACCC GGCGCTCCTC GACGCCGCAC TGCACCCCGC ACTCCTCGGC

20751 GAGCGGCCGG CCGAGGAGGA CAAGGTGTGG CTGCCGTTCA CGCTGACCGG

20801 AGTGACCCTG TGGGCCACCG GTGCCACCTC TGTACGCGTC CGTCTCACCC

20851 CGCTGGACGA CGACCCCGAC GCGTCGGCGG ACGGGCGGGC CTGGCGGGTC

20901 GGCGTGAGCG ACCCGACCGG CGCGGAGGTG CTGACCTGCG AGGCCCTGGT

20951 CGCGGTGGCG GCGGGCCGCC GCGAGCTGCG GGCCGCGGGG GAGCGGGTGT

21001 CCGATCTGTA CGCGGTGGAG TGGGTGCCGG TGCCGGGCCC GGGGCCGGTG

21051 GGTGAGGGTG CTGACTTCTC GGGCTGGGCC GGTCTGGGGG AGTGCGGGGA

21101 GCGTTGGGAG TGCGTGGGGC GCGTGGAGCG CTGGTACGAG GACCTGGACG

21151 CTCTCGGCGC GGCTGTCGAG GGTGGGGCTT CGGTGCCCTC TGTCGTTCTC

21201 GCCACCGCGG CTGCCGCCCC TGGTGGAGCG GGCGACGGAG CCGCCGATGC

21251 GCTGAGCGCC GTGCGGTGGA CCGGCGCGCT CCTCGATCAG TGGCTCGCCG

21301 ACGCGCGGTT CGCCGACGCC CGGCTGGTGG TGATCACGTC CGGCGCGGTC

21351 GCCACGGGTG ACGATTTCCT TCCCGACCCG GCCGCCGCGG CGGTACGAGG

21401 ACTGGTCGAG CAGGCGCAGG TCAGGCACCC CGGCCGCATC CTCCTCGTCG

21451 ACACGGAAGC CGGGGCCGGG CTCGGGGTCG GCGCCGGAGT CGATGACGCG

21501 CTCCTGGAAC AGGCCGTGGC CATGGCTCTC GGCGCCGACG AACCGCAACT

21551 CGCCCTGCGC GCGGGGCGGG TCCTGGCGCC CCGCCTCACC GCACCCCAGG

21601 ATGCGGCCGT CACCGAAGCG GCGCGACCGC TCGACCCGGA CGGCACCGTA

21651 CTCATCACAG GGCCGGCCGG TGCTCCGGTG GCCGACCTCG CCGAACACCT

21701 CGTACGCACC GGGCAGTGCA GGCATCTGCT GCTCCTGCCT GGAGACGGTG

21751 AACTGGAGGA AATGGCCGAG GAGTTGCGGG GCCTCGGCGC CACCGTGGAC

21801 CTGAGTACCG CCGACCCGGC GGACCCGACC GCCCTCGCCG AAGTGGTCGC

21851 CGCCGTCGAG GGGGACCATC CTCTTACGGG GGTCATCCAC GCCACCGGAG

21901 TCGTGGACGC GTTCGATCCC GGCGACTCGG CGAGCGACTT GATGATCGAC

21951 TCGGCGAGCG ATTCGTTCGC CGAGGCATGG TCGTCGAGGG CGGGCGTCAC

22001 CGCCGCACTG CACACCGCGA CCGCCCACCT TCCCCTGGAC CTGTTCGCCG

22051 TCCTGTCCCC GGCGGGCGCG GACCTGGGCA TTGCCCGGTC GGCGGCCGCC

22101 GCGGGCGCCG ACGCCTTCAG CGCGGCACTC GCCCTGCGCC GGCACACGAC

22151 CGTCACGACG GACACGACAG CCCCGCCGCG CACGACAGCC CCGCCGCGAA

22201 CGACAGCCTC GCCGCGCACG ACAGCCCTGT CGTCGTCGCG CACGACGGGC

22251 GTGGCCCTCG CCTACGGGCC GCCCACCGCG CCGAGGCCCG GCATCAAGGG

22301 GACGGCGCCC GGTCGGATCC CCGTGCTGCT CGACGCCGCT CGCGCTCACG

22351 GGGGCGGTTC GCCCCTGCTC GGGGCCCGCT TGGCCGCGCG TGCCCTGGCC

22401 GCCGAGTCCG CCGCCGAGGG CGTCGCCGGC CTGCCCGCGC CGCTGCGCGC

22451 GCTGGCAGTG GCCGCAGCCG CGGCCGGAGC ACCGACCCGG CGCACCGCCG

22501 CCGACCGCAA GCCCCCCGCG GACTGGCCGG CCCGACTGGC CCCCCTGTCC

22551 GCCCCCGAAC AACTCCGTCT GCTCATCGAC GCCGTACGCA CCCACGCCGC

22601 CGCGGTCCTC GGCCGCACCG ACCCGGAAGC GCTGCGCGGG GACGCCACCT

22651 TCAAGCAGCT CGGCCTTGAC TCGCTGACCG CCGTGGAGCT GCGCAACCGG

22701 CTCGTGGAGG ACACCGGTCT GCGCCTGCCC ACCGCCCTCG TCTTTCGCTA

22751 CCCGACCCCC GCGGCGATCG CCGCGCACCT CCGCGAGCGG CTGACCAGCC

22801 CGAGCGAGAC GACCGCCACA CAGAGGTCCG GAGGGCAGAC GCCCGCAGCG

22851 GGGCAGGCGT CGTCCGCGCT CGCCCCCGGC GGATCGGCCG CCGGACCGCC

22901 CGCCGCAGAC ACCGTGCTGA GCGACCTGAC CCGCATGGAG AACACCCTCT

22951 CCGTGCTCGC CGCCCAGCTG CCCCACACCG AGACGGGTGA GATCACCACC

23001 CGGCTCGAAG CGCTCCTCAC GCGCTGGAAG ACCACGAACG CCACGGCGAA

23051 CGACAGCGGC GACGGCAACG GCGGCGATGA CGACGCCGCC GAACGCCTCA

23101 AGGCCGCGTC CGCCGACCAG ATCTTCGACT TCATCGACAA CGAGCTTGGT

23151 GTCGGGCACG GCACCTCGCG CGTGACCCCC ACTCCGAAGG CCGGGTGACC

23201 GCACATGGCG AGTGAAGAGC AACTGGTCGA ATATCTGCGC AGGGTGACCA

23251 CCGAGCTCCA TGACACGCGT CGGCGCCTGG TGCAGGAGGA GGACCGCAGG

23301 CAGGAACCGG TGGCCCTGGT CGGCATGGCC TGCCGCTTCC CGGGCGGCGT

23351 GGCCTCACCG GAGGACCTCT GGGACCTGGT CGCCGCGGGC AAGGACGCCA

23401 TCGAGGACTT TCCCACCGAC CGGGGCTGGG ACCTGGAGGC GCTCTACGAC

23451 CCGGACCCGG CCGCGTACGG GACCAGCTAT GTCCGCCACG GCGGGTTCGT

23501 GGACGACGCG GGCTCCTTCG ACGCCGACTT CTTCGGCATC AGCCCGCGAG

23551 AAGCCCTGGC GATGGACCCG CAGCAGCGGC TGATGCTGGA GACGTCCTGG

23601 GAGCTGTTCG AGCGCGCCGG CATCGAACCC GTCTCCCTCA AGGGCAGCCG

23651 TACGGGCGTC TACGCCGGGG TGTCCAGCGA GGACTACATG TCCCAACTGC

23701 CCCGCATCCC CGAGGGGTTC GAGGGGCACG CCACCACCGG CAGCCTCACC

23751 AGCGTCATCT CGGGCCGGGT CGCGTACAAC TACGGCCTCG AAGGCCCGGC

23801 CGTCACCGTC GACACAGCCT GTTCCGCCTC GCTCGTCGCC ATCCACCTGG

23851 CGAGCCAGGC GCTGCGCCAG CGTGAGTGCG ACCTCGCCCT CGCGGGCGGT

23901 GTGCTCGTAC TGTCCAGCCC GCTCATGTTC ACCGAGTTCT GCCGCCAGCG

23951 GGGCCTTGCT CCCGACGGCC GCTGCAAGCC GTTCGCCGCC GCGGCGGACG

24001 GCACCGGCTT CTCGGAGGGC ATCGGTCTGC TCCTCCTGGA GCGCCTGTCC

24051 GACGCCGCGC GCAACGGCCA CAAGGTGCTC GCGGTGATCC GCGGCTCCGC

24101 CGTCAACCAG GACGGCGCGA GCAACGGCCT GACCGCCCCC AACGACGCCG

24151 CGCAGGAACA GGTCATCCGC GCCGCCCTCG ACAACGCCCG CCTCACCCCG

24201 TCCGAGGTGG ACGCCGTCGA GGCGCACGGC ACCGGCACCA AACTGGGCGA

24251 CCCCATCGAG GCCGGAGCGC TGCTCGCCAC CTACGGGCAA CACCGCGCCC

24301 GGCCCCTCCT CCTCGGCTCC CTCAAGTCCA ACATCGGCCA CACCCACGCC

24351 ACCGCGGGCG TCGCCGGTGT CATCAAGACC GTCATGGCGA TCCGCAACGG

24401 TCTGCTCCCC GCCACCCTCC ACGTCGAGGA ACTGAGCCCG CACGTCGACT

24451 GGGACGCGGG CGCGGTCGAG GTCGTCACGG AGCCCACCCC GTGGCCCGAG

24501 ACCGGCCACC CCCGGCGCGC GGGCGTCTCC GCGTTCGGGA TCTCCGGGAC

24551 GAATGCGCAC TTGATCCTGG AGGAGGCGCC GCCGGAGGAG GATGTGCCCG

24601 CCCCCGTGGT TGTGGAGTCG GGCGGGGTCG TTCCGTGGGT GGTGTCCGGG

24651 CGGACGCCGG AGGCGCTGCG TGAACAGGCC CGGCGACTCG GCGAGTTCGT

24701 GGCAGGCGAC ACGGACGCAC TGCCGAACGA GGTCGGCTGG TCCTTGGCCA

24751 CGACCCGGTC GGTGTTCGAG CACCGGGCTG TGGTCGTGGG GCGTGACCGG

24801 GATGCGTTGA CGGCTGGCCT GGGGGCGTTG GCTGCGGGTG AGGCTTCGGC

24851 GGGTGTGGTG GCCGGGGTGG CCGGTGATGT GGGTCCTGGG CCGGTGTTGG

24901 TGTTTCCGGG GCAGGGGGCG CAGTGGGTGG GCATGGGTGC CCAGCTGTTG

24951 GACGAGTCTG CGGTGTTCGC GGCGCGGATC GCGGAGTGTG AGCGGGCCCT

25001 GTCGGCGCAT GTGGACTGGT CGCTGAGTGC GGTGTTGCGC GGGGACGGGA

25051 GTGAGCTGTC CCGGGTGGAA GTGGTGCAGC CGGTGCTGTG GGCGGTGATG

25101 GTCTCGCTGG CTGCGGTGTG GGCGGATTAC GGGGTCACTC CGGCTGCCGT

25151 GATCGGGCAC TCGCAGGGTG AGATGGCTGC CGCGTGTGTG GCGGGGGCGC

25201 TGTCGCTGGA GGATGCGGCG CGGATCGTAG CGGTACGCAG TGACGCGCTT

25251 CGTCAGCTGC AAGGGCACGG CGACATGGCC TCGCTCAGCA CCGGTGCCGA

25301 GCAGGCCGCT GAGCTGATCG GTGACCGGCC GGGCGTGGTC GTCGCGGCGG

25351 TCAATGGGCC GTCGTCTACG GTGATTTCAG GGCCGCCGGA GCATGTGGCA

25401 GCCGTGGTCG CGGATGCGGA GGCACGTGGT CTGCGCGCCC GTGTCATCGA

25451 CGTCGGCTAT GCCTCGCATG GCCCCCAGAT CGACCAGCTC CACGATCTGC

25501 TGACCGAACG CCTGGCCGAC ATCCGGCCCA CGAACACGGA CGTGGCCTTC

25551 TATTCGACGG TCACCGCCGA GCGCCTGACG GACACCACGG CCCTGGACAC

25601 GGATTACTGG GTCACCAACC TCCGTCAGCC CGTCCGGTTC GCCGACACCA

25651 TCGAAGCCCT TCTCGCGGAC GGCTACCGCC TGTTCATCGA GGCCAGCGCC

25701 CACCCCGTGC TGGGCCTGGG CATGGAGGAG ACCATCGAGC AGGCGGACAT

25751 GCCCGCCACC GTCGTCCCCA CCCTCCGCCG CGACCACGGC GACACCACCC

25801 AGCTCACCCG CGCCGCCGCC CACGCCTTCA CCGCCGGCGC CGATGTCGAC

25851 TGGCGGCGCT GGTTCCCGGC CGACCCCGCC CCCCGCACGA TCGATCTCCC

25901 CACCTACGCC TTCCAGCGCC GCCGCTACTG GCTGGCCGAC ACAGTGAAGC

25951 GGGACAGCGG ATGGGACCCG GCCGGGTCGG GGCATGCCCA GTTGCCGACC

26001 GCGGTCGCCC TCGCCGACGG GGGAGTGGTG CTGAACGGCC GGGTGTCCGC

26051 CGAGCGCGGT GGCTGGCTGG GCGGGCATGT GGTGGCGGGG ACGGTTCTGG

26101 TGCCGGGTGC GGCGTTGGTG GAGTGGGTGT TGCGGGCCGG TGATGAGGCG

26151 GGTTGCCCCT CGCTTGAGGA GTTGACGCTC CAGGCGCCGT TGGTGTTGCC

26201 CGAGTCGGGT GGGTTGCAGG TTCAGGTGGT CGTGGGTGCG GCTGATGAGC

26251 AGGGCGGCCG TCGTGACGTA CATGTGTATT CGAGGTCTGA GCAGGACGCG

26301 TCGGCGGTGT GGCAGTGCCA TGCCGTCGGT GAGCTCGGGC GCGCGTCGGT

26351 GGCGCGGCCG GTGCGGCAGG CCGGGCAGTG GCCTCCGGCG GGGGCCGAGC

26401 CGGTGGAGGT GGGCGGCTTC TACGAGGGCG TCGCGGCCGC CGGTTACGAG

26451 TACGGTCCGG CGTTCCGTGG GCTGCGCGCG ATGTGGCGGC ACGGTGATGA

26501 CCTCCTTGCG GAGGTCGAGC TGCCGGAGGA GGCCGGTTCG CCGGCCGGTT

26551 TCGGCATCCA CCCGGCGCTG CTGGACGCCG CCCTGCACCC GCTGCTCGCA

26601 CAGCGGAGCC GGGACGGGGC CGGGGCGGGG GCCCACGGCG GGCAGGTGCT

26651 GCTGCCTTTC AGCTGGAGCG GTGTTTCCCT GTGGGCCAGC GAGGCCACCA

26701 CTGTGCGGGT GCGGCTCACC GGGCTGGGAG GAGGGGACGA CGAGACGGTG

26751 TCCCTGACGG TAACCGACCC CGCCGGTGGC CCCGTGGTGG ACGTGGCAGA

26801 GCTGCGGTTG CGGTCGACGA GCGCCCGGCA GGTGCGGGGT TCGGCAGGCC

26851 CCGGCGCGGA CGGGCTCTAC GAGCTGCGGT GGACACCGTT GCCCGAGCCG

26901 CTTCCCGTAC CGGCCCCCGC GAACGGTCGC GATGTGGCCG CCGACCTGTC

26951 CGGATGCGCG GTGCTCGGCG AACTGGTCGC GGAACCGGGC CCGGGCATCG

27001 ACCTGGAGGG CTGCCCCTGC TACCCGGGCG TCGGCGCGCT CGCCGACAAC

27051 GCCTCCCCGC CCTCCATGAT CCTCGCCCCC GTGCACAGCG ACACCACAGG

27101 CGGCGACGGA CTCGCCCTGA CGGAACGGGT GTTGCGCGTC ATCCAGGACT

27151 TCCTGGCTGC ACCGAGTCTG GAACAGAAAC AGACGCGCCT GGCCTTCGTG

27201 ACCCGGGGCG CGGCGGACAC AGGTAGCACG ACGGGAGGCT CGGCTGCCCC

27251 GGCAGAGGCA GTCGACCCGG CGGTCGCGGC CGTATGGGGC CTAGTACGCA

27301 GCGCGCAGTC GGAGAACCCC GGCCGCTTCG TACTGCTGGA CACCGACGCG

27351 CCCCTCGACC AGGCGTCCGT TGCCCCTCTC GTGGACGCGG TGCGGTCTGC

27401 CGTGGAGGCG GACGAGCCCC AAGTCGCCCT GCGCGGGGGA CGGTTGCTCG

27451 TGCCCAGGTG GGCGCGGGCC GGCGAGCCCG TCGAGCTGGC CGGGCCGGCC

27501 GGAGCGCGGG CGTGGCGGCT GGTGGGCGGA GACTCCGGGA CGCTGGAGGC

27551 CGTCGTGGCG GAGGCTTGCG ACGACATTGT GCTGCGCCCG TTGGCGCCGG

27601 GCCAGGTCCG CGTCGCCGTC CATACGGCCG GGGTCAATTT CCGTGACGTC

27651 CTGATCGCCC TGGGCATGTA CCCGGACCCG GACGCGCTGC CCGGCACCGA

27701 GGCGGCCGGC GTGGTGACGG AGGTCGGGCC GGGCGTCACC CGTCTGTCGG

27751 TGGGCGACCG CGTGATGGGC ATGATGGACG GCGCCTTCGG CCCGTGGGCC

27801 GTCGCCGACG CGCGCATGCT GGCCCCGGTC CCGCCCGGCT GGGGCACCCG

27851 GCAGGCGGCC GCCGCTCCCG CCGCGTTCCT GACGGCTTGG TACGGGCTGG

27901 TGGAGCTGGC CGCTCTGAAG GCGGGCGAGC GTGTGTTGAT CCATGCCGCC

27951 ACGGGTGGTG TGGGGATGGC GGCGGTGCAG ATCGCCCGGC ATGTGGGTGC

28001 CGAGGTGTTC GCCACCGCGA GTCCGGGCAA GCACGCCGTG CTGGAGGAGA

28051 TGGGCATCGA CGCCGCCCAC CGCGCCTCGT CGCGCGACCT CGCCTTCGAG

28101 GACGCCTTCC GGCAGGCCAC CGACGGCCGT GGCGTGGACG TCGTCCTCAA

28151 CAGCCTCACC GGTGAACTGC TCGACGCGTC CCTGCGATTG CTCGGCGACG

28201 GCGGGCGCTT CGTGGAGATG GGCAAGAGCG ATCCGCGCGA CCCCGAGCTG

28251 GTCGCGCTGG AGCACCCCGG GGTGTCGTAC GAGGCCTTCG ACCTCGTCGC

28301 CGACGCCGGG CCCGAGCGGC TCGGGCTGAT GCTCGACAGG CTCGGCGAGC

28351 TCTTCGCCGG CGGATCACTG GTACCGCTGC CGGTCACCGC ATGGCCGCTG

28401 GGGCGGGCGC GAGAGGCGCT CCGCCACATG AGTCAGGCGA GGCACACCGG

28451 CAAGCTGGTG CTCGACGTGC CCGCGCCGCT CGACCCCGAC GGCACCGTCC

28501 TCGTCACCGG GGGTACCGGC ACCATCGGCG CCGCCGTGGC CGAACACCTG

28551 GCGCGTACCG GGGAGAGCAA GCACCTGCTC ATCGTCAGCC GCAGCGGGCC

28601 GGCCGCCCAC GGCGCCGAGG AACTTGTCTC TCGTATAGCC GAGTTCGGGG

28651 CCGAAGCCAC CTTCGTCGCT GCCGACGTGA GTGAGCCCGA CGCGGTCGCC

28701 GCCCTGATCG AAGGGATCGA TCCGGCCCAT CCGCTGACCG GTGTCGTGCA

28751 TGCCGCCGGA GTACTCGACA ACGCTCTGAT CGGCTCCCAG ACCACCGAAA

28801 GCCTCACCCG CGTATGGGCG GCGAAGGCCG CCGCCGCGCA GCAACTCCAC

28851 GAGGCCACGA GGGAGTCGAG GCTGGGACTG TTCGTGATGT TCTCCTCCTT

28901 CGCCTCCACC ATGGGCACCC CAGGGCAGGC CAACTACTCC GCCGCCAACG

28951 CCTATTGCGA CGCGCTGGCC GCTCTCCGAC GCGCGGAGGG GCTCGCCGGC

29001 CTGTCCGTGG CGTGGGGGTT GTGGGAGGCC ACCAGCGGCC TGACCGGGAC

29051 GTTGTCGGCG GCCGACCGGG CCCGCATCGA CCGGTACGGC ATCAGGCCGA

29101 CCAGCGCGGC ACGCGGCTGC GCCCTGCTGG CAGCGGCACG CGCCCACGGG

29151 CGCCCCGACC TGCTCGCCAT GGACCTGGAC GCCCGCGTAC CCGCCGCGTC

29201 CGACGCTCCG GTCCCCGCCG TGCTGCGCAC TCTGGCGGCC GCCGGAGCGC

29251 CCGCCACCGC CCGTCCCACC GCGGCGGCGG CCGCTGACGG GGCGACGGAC

29301 TGGTCCGGCA GGCTCGCCGG CCTCACCGAG GAGGCACGGC TCGAACTCCT

29351 CACCGAGTTG GTGTGCACCC ACGCGGCAGG GGTGCTCGGG CACGCCGACG

29401 CGGGCGCGGT CCAGGTGGAC GCGCCGTTCA AGGAACTCGG CTTCGACTCG

29451 CTGACCGCCG TCGAACTGCG CAACCGGATC GCCGCCGCGA CCGGCCTGAA

29501 ACTGCCCGCC GCCCTCGTCT TCGACTACCC GCAGGCTCGC GTTCTCGCCG

29551 CCCACCTGGC CGAACGGCTC GTCCCGGAGG GCGCGGGGGC CATGGGCGGT

29601 GTGAGCGGTG CGGAGGGCGT GAGGGACGCG TACGGGGCAG GCGGTCCGGG

29651 CGGCGACATG ACCGCCCAGG TCTTGCTGGA GGTGGCCCGC GTCGAGCACA

29701 CCCTGTCCGC CGCCGTCCCG CACGGCCTGG ACCGGGCGGC CGTCGCGGCC

29751 CGCCTGGAGG CGCTGCTCGC CCGCTGCACG GCGACGACGG CGGCCACGGG

29801 GGCCGCGGGA GCCGCGGTGG AGGGTGACGG CGACAGCGAC GGCGACGGCG

```
29851  CCGTGGATCA GCTGGAGACG GCCACCGCCG AGCAAGTACT GGACTTCATC
29901  GACAACGAAC TCGGGGTGTG AGCCGCGTGC CGGCCGCACA CCAGGCGATC
29951  ACGGGCGGGG AGCTGCAGCG CACATGGTGA GCGAAGAGAA ACTGGTCGAC
30001  TACCTCAAGC GTGTCTCCGC GGACCTGCAC GCCACCCGGC AGCGGCTGCG
30051  CGAGGCGGAG GAGCGCGGCC AGGAACCCGT GGCCGTGGTG GAGGCCGCCT
30101  GCCGCTACCC CGGCGGCATC CGCACCCCCG AAGACCTGTG GGACCTGGTC
30151  GCCGCGGGCG GCAACGCCCT GGGCGCCTTC CCCGACAACC GCGGCTGGGA
30201  CCTGCGACGC CTCTTCCACC CCGACCCCGA CCACCCCGGG ACGACCTACG
30251  CCCGCGAGGG CGGCTTCCTC CACGACGCCG ACCTGTTCGA CCCGGAGTTC
30301  TTCGGCATCA GCCCCCGCGA GGCCGCGGTC CTCGACCCGC AGCAGCGACT
30351  GCTCCTGGAG TGCGCCTGGG AGGCACTGGA GCGCGCGGGC ATCGACCCGC
30401  GGTCCCTCCA GGGCAGCCGT ACCGGCGTGT ACGCGGGTGC CGCCCTGCCC
30451  GGCTTCGGCA CCCCGCACAT CGACCCCGCC GCCGAGGGCC ACCTGGTCAC
30501  CGGCAGCGCC CCGAGCGTCC TCTCGGGCCG GCTCGCCTAC ACCTTCGGCC
30551  TCGAAGGGCC CGCGGTGACG ATCGACACCG CCTGCTCGTC GTCGCTCGTC
30601  GCCGTGCACC TGGCGGCCCA CGCGCTGCGG CAGCGCGAGT GCGATCTGGC
30651  GCTCGCGGGC GGTGTCACCG TCATGACCAC CCCGTACGTG TTCACCGAGT
30701  TCTCGCGCCA GCGCGGCCTG GCCGCCGACG GCCGGTGCAA GCCCTTCGCG
30751  GCCGCCGCGG ACGGCACGGC CTTCTCCGAG GGCGCCGGAC TCCTCGTACT
30801  GGAACGCCTC TCCGACGCCC GCCGGGCCGG CCACCGGGTG CTCGCCGTCA
30851  TCCGCGGCTC GGCCGTCAAC CAGGATGGCG CGAGCAACGG CCTCACCGCC
30901  CCCAACGGCC CCGCCCAGCA GCGCGTGATC CGCGCCGCCC TCGCCGGGGC
30951  GCGGCTCTCG CCCGCGGAGG TGGACGCGGT CGAGGCGCAC GGCACCGGCA
31001  CCCGGCTGGG CGACCCCATC GAGGCCGACG CGCTCCTCGC CACCTACGGT
31051  CAGGAGCGCC ACGGGGGCCG GCCGCTGTGG CTCGGCTCGG TGAAATCCAA
31101  CATCGGCCAC ACGCAGGGCG CGGCCGGTGC CGCGGGCCTG ATCAAGATGG
```

```
31151  TCCAGGCACT GCGGCACGAG ACGCTGCCCG CCACGTTGTA CGCCGACGAG
31201  CCCACCCCGC ACGCCGACTG GGAGTCGGGC GCGGTGCGCC TGCTCAGCGC
31251  GCCGGTCGCC TGGCCGCGCG GGGAGCACGG GGAGCACACC CGCAGGGCCG
31301  GCATCTCCTC CTTCGGCATC TCCGGCACGA ACGCCCACCT CATCCTGGAG
31351  GAGGCGCCCG CGGCCGACGC CGAAGGAGCG GGTGGCGACG GCGATGGCGA
31401  CGGGGGAGGG GTGCGGCCGG TGGTGCGGGT CGGCGCCACG GGCCCCCGCG
31451  AAGAGCAGGG CCAAGGACAG GGCCAAGAGC AGCACCAACA GCAACGTCAG
31501  CAGCGGCAGC GGTCGTCGAT GATGCCGACG CCGCACCTCC CGTGGCTGCT
31551  GTCCGCCCGC AGCCCCGCCG CGCTCCGCGC CCAGGCCGAC GCGCTGGCGA
31601  ACCATGTCGC CCACGCGGAC CACTCCATCG CCGACATCGG CGGCACACTG
31651  CTGCGCCGCA CCCTGTTCGA GCACCGGGCG GTCGTCCTCG GAACCGACCG
31701  TGATGAGCGT GCCGCAGCGC TTGCCGCCCT CGCGGCAGGA CGCGCACACC
31751  CCGCGCTGAC CCGGGCCGCA GGGCCGGCGA GGAACGGCGG CACCGCCTTC
31801  CTGTTCACCG GCCAGGGAAG CCAACGCCCA GGCATGGGCA GGCAGTTGTA
31851  CGACACCTTC GACGTCTTCG CCGAGTCGCT CGACGAGACC TGCGCCCGGC
31901  TCGACCCCCT GCTCGAACAG CCGCTGAAGC CCGTCCTGTT CGCCCCCGCC
31951  GACACCGCGC AGGCCGCCGT GCTGCACGGG ACCGGCATGA CGCAGGCCGC
32001  GCTGTTCGCC CTCGAAGTCG CCCTGTACCG CCAGGTCACC TCCTTCGGGA
32051  TCGCCCCCAG CCACCTGACC GGGCACTCCG TCGGCGAGAT CGCCGCCGCC
32101  CACGTCGCCG GGGTGTTCTC CCTGGCGGAC GCCTGCACGC TGGTCGCGGC
32151  CCGGGGCCGC CTCATGCAGG CCCTGCCCGC AGGTGGCGCC ATGCTCGCCG
32201  TCCAGGCGGC CGAGGACGAC GTACTGCCGC TGCTCGCCGG GCAGGAGGAA
32251  CGTCTCTCCC TCGCCGCCGT CAACGGCCCC ACCGCCGTCG TCGTGTCCGG
32301  TGAGGCCGCT GCCGTCGGGG AGGTGGAGAA GGCGCTGCGC GGGCGCGGAC
32351  TGAAGACCAA GCGGCTCAAC GTCAGTCACG CCTTCCACTC GCCGCTCATC
32401  GAGCCGATGC TCGACGACTT CCGCGAAGTG GCCCGCGGGC TGACCTTCCA
```

```
32451  CGCGCCGACG CTGCCCGTCG TCTCCAACCT CACCGGCCGC CTCGCCGACG
32501  CGGAGCTGAT GGCCGACGCC GAGTACTGGG TGCGGCACGT ACGCCGGCCG
32551  GTGCGGTTCC ACGACGGGCT GCGCGCTCTC AGCGAGCAAG GCGTCGTGCG
32601  CTACCTGGAG TTGGGGCCCG ACCCGGTCCT CGCCACCATG GTCCAGGACG
32651  GTCTCCCGGC CCCGGCGGAG GGAGAGGAGC CCGAGCCGGT CGTCGCCGCG
32701  GCGCTGCGCT CCAAGCACGA CGAGGGACGC ACCCTGCTGG GTGCCGTCGC
32751  CGCGCTCCAC ACCGACGGAC AGCCGGCCGA CCTCACCGCC CTCTTCCCCG
32801  CCGACGCCGG GCAAGTGCCG CTCCCCACCT ACCGGTTCCA GCGGCGACGG
32851  TACTGGCGCG TCGCGCCCGA CGCCGCCGCG CCGGCCCGCG CCGCCGGCCT
32901  CCAGGAGACC GGCCACCCGC TGCTGCCCGC CGTCATCCGG CAGGCCGACG
32951  GCGGCATCCT GCTCGCGGGA CGCCTGTCCC TGCGTACGCA TCCATGGCTC
33001  GCCGACCACA CCATCGCGGG CGGCGTCCCG CTGCCCGCCA CCGCCTTCGT
33051  CGAACTCGCC CTGCTCGCAG GGCGGCACGC CGCCTGCGAC ACGATCGACG
33101  ATCTGACGCT GGAGACGCCG CTGCTGCTCG ACGACACCGG TACCGGTGTC
33151  GGGGCGGCTG TGGGCGCGGG CGCCGATGCC CTCGTCGATG CCATAGAAGT
33201  GCAGCTTGCC CTCGGCGCTC CCGACGGTTC CGGCCGCCGT GCTCTCACCG
33251  TCCACTCCCG TCCTGCCGAC GATGCGGCTG ACGACGGCGA CGCGGCCGAC
33301  GCGGCCGATG CGGCAGGCCG GGGAGGCCCG GGCGGCTCGG GTGACCTGGG
33351  CGATCCTGGC GATCCGGGCG ATCTGGGCGA CGGCGGGGGC TCCCGCGGCT
33401  GGCGCCGTCA CGCCACCGGC ATCCTCAGCG CCGGCCCGGC CGCCGAACCG
33451  GCCGCCCCCG ACGCCGCTCC CTGGCCGCCC GCCGACGCCA CCGCCCTCGA
33501  CGTCGACGCG CTGTACGCCC GGCTCGACGC GCAGGGCTAC AGCTACGGGC
33551  CCGCCTTCCG GGCCGTCCAC GCCGCCTGGC GGCACGGCGA CGACCTCTAC
33601  GCCGATGTCC GCCTCGCCGA CGAACAGCGC GCTGAAGCCG ACGCGTTCGC
33651  CCTCCACCCG GCCCTGCTCG ACGCCGCCCT GCATGCCGTC GACGAGCTGT
33701  ACCGCGGCAG TGAGGGGCGG GGGCAGGAGC AGGGGCAGGG TGGTCAGGAG
```

33751 CCGGAGCAGG GCCGTGGCGA CGCGGACGCC CCGGTACGGC TGCCGTTCTC

33801 CTTCAGCGAC ATACGCCACC ACGCCACCGG GGCCACACGG CTGTGGGTCC

33851 GCCTCAGCCC CCAGGGCGAC GATCGGCTGC GGCTGTCCCT GACCGACGGC

33901 GAGGGCGGGC AGGTCGCGAC AGTCGACGCC CTCCAACTGC GGTTGATCCC

33951 CGCCGACCGG TGGCGCGCGG CCCGCCCCAC CACAGCCGCC CCCCTGTACC

34001 ACCTGGACTG GCACGAGCTG CCGTTGCCCG AGCCGGCCGA GACGGACCCG

34051 GCCGCCCACT CCTGGGCTGT GCTCGGAGCG CACGACGCGG GCCTCGCTCC

34101 CGCCGCGCAC TACCCGGACC TGGCGGCCCT GAAAGCCGCC GTCGAGGCCG

34151 GCGAGCCCGT GCCGGACATC GTCTTCGCAC CGTTCCCCGC GCAGGGGACG

34201 GAGACCGATG TCCCGGCTCA GGTACGAGCC CACGCCCGGC ACGCCCTGGA

34251 GCTGCTGCGC GACTGGCTCA CCACGGAAGC TTTCGCCGCC GCCCGCCTCG

34301 TCGTCCTCAC GACCGGTGCG GTCACCGCCC GCCCAGAGGA CGGGCCCGCC

34351 GACCTGGCCA CCGCACCTGT ATGGGGCCTG GTCCGAGCCG CCCAGGCCGA

34401 ACAACCCGAC CATGTCGTCC TGGTGGACAT CGACAAGGAC ATCGATAAGG

34451 ACACCGACGA GGAGACCGAC CAGGCCACCG ACGCGGGCAC CGCATCGCGC

34501 CACGCTCTGC CCGCCGCCTT GGCCGCGGCG GCCGCCCAAG CCGAGACACA

34551 GCTCGCCCTG CGCGCGGGCA CCGTGCTCGT GCCGCGCCTC GCCGTCGTCC

34601 CGCCCCGGAC CGACACCCCA GCGCTGCACG CCACCGCCCC GGAGAGCACC

34651 ACGGACACTG TGGACTCCAC GGGCATCGCG GGCGCTGCGG AATCCGGCGG

34701 CACCGTCCTG ATCACCGGCG GAACCGGCGG CCTCGGGCAG GCCGTCGCCC

34751 GTCACCTCGC CGCCGCGCAT GGCGCCCGCC ACCTGCTCCT CGTCAGCCGC

34801 AGGGGCGACG CCGCCGAGGG CGTCGCCGAG TTGCGCGCCG ACCTCGCGGA

34851 CGACGGCGTC GACGTACGCG TCGCCGCCTG CGACATCACC GACCGCGACG

34901 CGCTGGCCGG GCTCCTCGCG GACATCCCCG CCGCGCACCC GCTCACCGCG

34951 GTCGTGCACA CCGCGGGCGT CATCGACGAC AGCCTCATCA CGGCGATGAC

35001 CCCCGAGCGG CTCGACGCCG TCCTCGCACC CAAGGCCGAC GCGGCCTGGC

35051 ACCTGCACGA ACTCACCCGC GACAAGGACC TGTCGGCCTT CGTCCTCTTC

35101 TCCTCGGGCG CCTCCGTCCT CGGCAACGGC GGCCAGGCCA ACTACGCGGC

35151 CGCCAACACC TTCCTCAACA CCCTCGCCGA ACACCGCCGC GCGGCCGGCC

35201 TCGCCGCCAC CTCCGTGGCC TGGGGCCTGT GGGAGTCCGC GTCCGGCGGC

35251 ATGGCCGCCC GGCTCGGCGA CGCCGACCGC GCCCGCATCC ACCGCACCGG

35301 CGTGACGGGC CTGACCGACG AGCAGGCCCT GGCCCTCTTC GACGCGGCCC

35351 TGACCGCCGA GCACCCCACG GTCCTCGCCA CCCGCTTCGA CCGCGCCGTG

35401 CTGCGCGGCC AGGCCGCCGC CCGCACCCTG CAGCCCGCCC TGCGCGGCCT

35451 GGTACGCACT CCGCGCCCCA CCGCGTCCGC CGGGGCCATC GGGTCCACCG

35501 CAGCCACCGG GTCCGCCACG GACGAGAACG CGCCCTCCTC GTGGGCCGCC

35551 CGGCTCGCCC GGCTGTCCGC CGCCGACCGC GACCGCGCCC TCAACGAACT

35601 CATTCGCGAG CAGATCGCGA CCGTCCTGGC ACACCCCTCA CCCGACACCA

35651 TCGAACTGGG CCGCGCCTTC CAGGAGTTGG GCTTCGACTC GCTCACCGCC

35701 CTGGAACTCC GCAACCGCCT CTCCACGGCC ACCGGCATCC GGCTGCCCGC

35751 CACCCTCGTC TTCGACCACC CGAGCCCCAC CGCCCTCGTA CGCCATCTCC

35801 ACAGCCATCT CCCCGACGAG GCCCAGCACA CGTCCCCGAC CGCCCCCGGC

35851 GCCTCTGCGG AGGGCACCGC CGCCACGGCC ACCGGCATCG ACGACGACCC

35901 GATCGCCATC GTCGGCATGG CGTGCCGCTA CCCGGGCGGC GTGACCTCGC

35951 CCGAGCAGCT GTGGCAGCTC GTGGCCACCG GCACCGACGC CATCGGCCCG

36001 TTCCCCGAGG ACCGCGGCTG GGACACGGCC GGACTGTTCG ATCCCGACCC

36051 CGACCAGGTC GGCCACAGCT ACACCCGCGA AGGCGGCTTC CTCTACGACG

36101 CCGCCCGCTT CGACGCGGGC TTCTTCGGCA TCAGCCCGCG CGAGGCCGCC

36151 GCCACCGACC CGCAGCAGCG CCTGCTCCTG GAAACCGCCT GGCAGGCGTT

36201 CGAACACGCG GGCATCGACC CCGCCGCCCT GCGCGGCACC CCGTGCGGCG

36251 TCATCACCGG AATCATGTAC GACGACTACG GATCCCGCTT CCTCGCGCGC

36301 AAACCGGACG GCTTCGAGGG CCGCATCATG ACCGGCAGCA CGCCGAGCGT

```
36351  GGCCTCCGGC CGGGTCGCGT ACACCTTCGG CCTGGAGGGC CCCGCCATCA
36401  CGGTGGACAC CGCGTGCTCC TCCTCGCTGG TCGCGATGCA CCTGGCGGCG
36451  CAGGCGCTGC GGCAGGGCGA GTGCGAACTG GCCCTGGCCG GGGGTGTGAC
36501  CGTGATGGCC ACCCCGAACA CCTTCGTGGA GTTCTCCCGC CAGCGCGGCC
36551  TGGCCCCCGA CGGCCGCTGC AAGCCGTTCG CCGCCGCGGC GGACGGCACC
36601  GGCTGGGGCG AGGGCGCCGG ACTCGTCGTC CTGGAGCGCC TCTCCGACGC
36651  GCGCCGCAAG GGACACCGCG TCCTCGCCCT GCTGCGCGGT TCGGCCGTGA
36701  ACCAGGACGG CGCGAGCAAC GGCATGACCG CCCCGAACGG TCCCTCGCAG
36751  GAACGGGTCA TCCGCACCGC CCTGGCCGGC GCGGGCCGTG GTCCCGAGGA
36801  CATCGACGTG GTGGAGGCGC ACGGCACCGG CACCACGCTC GGCGACCCGA
36851  TCGAGGCGCA GGCCCTGCTC GCCACGTACG GGCAGGGGCG CCCGGAGGAC
36901  CGCCCGCTCT GGCTCGGCTC GGTGAAGTCG AACATCGGCC ACACGCAGGC
36951  CGCCGCCGGT GTCGCGGGCG TCATCAAGAT GGTCATGGCA CTGCGCCACG
37001  AGCAACTGCC CACGACCCTG CACGCCGACG AGCCGACCCC CCACGTGCAA
37051  TGGGACGGCG GCGGCGTACG TCTCCTGACC GAACCGGTCC CGTGGTCGCG
37101  CGGCGAGCGC ACGCGGCGCG CCGGGGTGTC GTCCTTCGGG ATCTCCGGGA
37151  CGAACGCGCA CCTGATCCTG GAGGAGCCGC CGGAGGAGGA CCTGCCCGAG
37201  CCCGTGGCGG CGGAGCCGGG TGGGGTGGTG CCGTGGGTGG TGTCCGGGCG
37251  GACGCCGGAC GCGTTGCGTG AACAGGCGCG GCGGCTCGGC GAGTTTGTCG
37301  TCGGTGCCGG GGATGTGTCG GCAGCCGAGG TGGGATGGTC ACTGGCCACG
37351  ACGCGGTCGG TGTTCGAGCA CCGGGCCGTG GTGGCGGGCC GGGACCGGGA
37401  CGATCTGGTT GCCGGGATGC AGGCGCTGGC GGCAGGGGAG ACGCCGACAG
37451  ATGTCGTGTC CGGTGCGGCG GCTTCCTCCG GTGCCGGGCC GGTGTTGGTG
37501  TTCCCGGGGC AGGGGTCGCA GTGGGTGGGC ATGGGTGCCC AGCTCCTTGA
37551  CGAGTCCCCC GTCTTCGCGG CGCGGATCGC GGAGTGTGAG CAGGCGCTGT
37601  CGGCGTACGT GGACTGGTCG CTGAGTGATG TCCTGCGCGG GGACGGGAGT
```

```
37651  GAGCTGTCCC GGGTCGAGGT CGTGCAGCCC GTGTTGTGGG CGGTAATGGT
37701  CTCGCTGGCT GCCGTCTGGG CGGATTACGG GGTCACTCCG GCCGCTGTGG
37751  TGGGGCATTC GCAGGGTGAG ATGGCTGCCG CGTGTGTGGC GGGGGCGCTG
37801  TCGCTGGAGG ATGCGGCGCG GATTGTGGCG GTACGCAGTG ACGCGCTTCG
37851  TCAGCTGCAA GGGCACGGCG ACATGGCCTC ACTCGGCACT GGTGCCGAGC
37901  AGGCCGCTGA GCTGATCGGT GATCGGCCGG GAGTGGTCGT CGCGGCAGTC
37951  AACGGGCCGT CGTCTACCGT GATTTCGGGG CCGCCGGAGC ATGTGGCCGC
38001  TGTGGTCGCG GAGGCGGAGG CACGTGGTCT GCGCGCCCGT GTGATCGACG
38051  TCGGGTATGC CTCGCACGGC CCCCAGATCG ACCAGCTCCA CGACCTCCTC
38101  ACCGAGGGCC TGGCTGACAT CCGGCCCGCG AACACGGACG TGGCCTTCTA
38151  TTCGACGGTC ACCGCCGAGC GCCTGACGGA CACCACAGCC CTGGATACGG
38201  ATTACTGGGT GACCAACCTC CGCCAGCCGG TCCGGTTCGC CGACACCATC
38251  GAAGCGCTTC TCGCGGACGG CTATCGCCTG TTCATCGAGG CCAGCGCGCA
38301  CCCGGTGTTG GGCCTGGGCA TGGAGGAGAC CATCGAGCAG GCGGACATCC
38351  CTGCCACGGT CGTCCCCACC CTGCGCCGCG ACCACGGCGA CACCACCCAG
38401  CTCACCCGCG CCGCCGCCCA CGCCTTCACC GCCGGCGCCG ATGTCGACTG
38451  GCGACGCTGG TTCCCGGCCG ACCCCACCCC CCGTACCGTC GACCTCCCCA
38501  CCTACGCCTT CCAGCACCAG CACTACTGGC TGGAGGAGCC CAGTGGGCTC
38551  ACCGGAGACG CCGCCGACCT CGGCATGGTG GCCGCCGGGC ATCCGCTGCT
38601  CGGTGCCTGT GTGGAACTCG CGGAGAGCGA CTCGTACTTG TTCACCGGGC
38651  GGCTCTCGCG CAGGGCTCCG TCCTGGCTGG CCGAACACGT GGTGGCGGGG
38701  ACGGTTCTGG TGCCGGGTGC GGCGTTGGTG GAGTGGGTGC TGCGGGCCGG
38751  CGATGAGGCG GGATGCCCGA CGATTGAGGA ACTGACGCTC CAGGCGCCGT
38801  TGGTGCTGCC CGAGTCGGGC GGGTTGCAGG TTCAGGTGGT CGTGGGTGCG
38851  ACCGATGAGC AGAGCGGCCG TCGTGACGTA CACGTGTATT CGAGGTCTGA
38901  GCAGGACGCG TCGGCGGTGT GGGTGTGCCA TGCCGTCGGT GTGGTGAGCT
```

38951 CCGAAATGCC AGAAGCGGCA GCCGAGTTGA GTGGGCAGTG GCCTCCTGCC

39001 GGGGCCGAAG CCGTGGATGT CGAGGACTTC TACGCGCGGG CCGCGGAGGC

39051 CGGATACGCC TACGGTCCGG CGTTCCAGGG GCTGCGGGCG CTGTGGCGGC

39101 ACGGGACGGA GCTGTTCGCC GAGGTGGTGC TGCCCGAACA GGCGGGTGGG

39151 CACGACGGTT TCGGCATCCA CCCGGCGCTG CTGGACGCCG CCCTGCATCC

39201 GCTGATGCTC CTCGACCGGC CCGCGGACGG GCAGATGTGG CTGCCGTTCG

39251 CGTGGAGCGG GGTGTCGCTG AACGCGGACC GGGCGACCCA CGTCCGTGTC

39301 CGGCTCTCCC CGCGGGGGGA GGCGGCCGAG CGTGACCTGC GGGTCGTCAT

39351 CGCCGACGCG ACCGGCGCGC CCGTCCTGAC GGTCGACGCC CTGACCCTGC

39401 GCGCGGCCGA TCCCGGCCGG CTGGGTGCGG CGGCCCGTGG CGGTGTCGAC

39451 GGCCTCTACA CCGTCGACTG GACCCCGCTG CCCCTGCCCC AGCCCCTTCC

39501 GCTGCCGCGG ACGGATGCAG GGGGGAGTGC CGACTGGGTC ATACTCTCGG

39551 ACAACTCCAG TGCAGCTCTG GCTGATGCCG TGTCGTCCGC GACGGCGGCA

39601 GGTGGCGGAG CGCCGTGGGC ATTGCTCGCT CCCGTGGGTG GCGGCTCTGC

39651 CGATGACGGG CTGCCGGTGG TGCGGCGGAC CCTCTCCCTC GTACAGGAGT

39701 TCCTGGCCGC CCCGGAGCTG ACCGAGTCCC GTCTCGTCAT CGTGACACGC

39751 GGTGCCGTGG CCACCGACGC CGATGGTGAC GTCGCGGCGT CCGCGGCAGC

39801 GGTATGGGGC CTGATCCGCA GCGCCCAGTC GGAGAACCCG GGCCGCTTCG

39851 TCCTGCTCGA CGTCGAGGAG GAGCACCTCC ACCCGGACGG CGGGGAACTG

39901 CCGTACGCCG CCCTGCGCCA CGCCGTAGAG GAGCTCGACG AGCCTCAACT

39951 TGCCCTCCGC AGCGGCAAAT TCCTCGTACC GCGCATGACG CCCGCCGCCG

40001 CCCCCGAGGA GCTCGTCCCG CCGGTCGGTA CGTCCGGCTG GCGCCTCGGC

40051 ACCTCCGGTA CGGCCACCCT GGAGAATCTG TCGGTGATCG ACGCTCCCGA

40101 GGCGTTCGCG CCGCTGGAGC CCGGGCAGGT GCGGATCTCC GTACGGGCGG

40151 CGGGCATGAA CTTCCGTGAC GTGCTGATCG CGTTGGGCAT GTATCCCGAC

40201 AAGGGCACGT TCGCGGGAAG CGAGGGCGCC GGACATGTGA CGGAGGTGGG

40251 ACCGGGCGTC ACTCATCTGT CGGTCGGTGA CCGGGTGATG GGTCTGTTCG

40301 AGGGCGCGTT CGCTCCGCTG GCCGTCGCGG ACGCCCGGAT GGTCGTCCCG

40351 ATTCCGGAGG GCTGGAGCTT CCAGGAGGCC GCGGCGGTGC CCGTGGTGTT

40401 CCTCACGGCC TGGTACGGCC TCGTGGACCT CGGCCGCCTC CGGGCGGGCG

40451 AATCGCTGCT CATCCACGCG GGCACCGGCG GAGTGGGCAT GGCCGCCACC

40501 CAGATCGCCC GCCACCTGGG CGCCGAGGTG TTCGCCACCG CGAGCCCCGC

40551 CAAGCACGGC GTGCTCGACG GCATGGGCAT CGACGCGGCC CACCGCGCCT

40601 CCTCCCGTGA CCTCGACTTC GAGGAGACCT TGCGGGCGGC GACGGGCGGG

40651 CGCGGCATGG ACGTCGTACT CAACAGTCTG GCCGGGGAGT TCACCGACGC

40701 CTCGCTGCGG CTGCTCGCCG AGGGCGGGCG CATGGTGGAC ATGGGCAAGA

40751 CCGACAAGCG CGACCCCGAC CGGGTCGCGG CCGAGCACGC GGGCGCGTGG

40801 TACCGGGCCT TCGACCTCGT GCCGCACGCG GGGCCCGACC GGATCGGGGA

40851 AATGCTGGCG GAGCTGGGCG AGTTGTTCGC CTCCGGCGCC CTGGCGCCGC

40901 TGCCCGTCCA GACCTGGCCG CTGGGCCGGG CGCGTGAGGC GTTCCGGTTC

40951 ATGAGCCAGG CGAAGCACAC CGGCAAGCTG GTGCTGGAGA TCCCGCCCGC

41001 CCTCGATCCG GACGGCACGG TGCTCATCAC CGGCGGCACC GGGGTCCTCG

41051 CCGCCGCGGT GGCCGAGCAT CTGGTGAGGG AGTGGGGCGT ACGACACCTG

41101 CTGCTGGCCG GGAGGCGCGG TTCCGAGGCG CCCGGGAGCA GTGAACTCGC

41151 CGAGGAACTG ACCGAGTTGG GGGCCGAGGT GACCTTTGCC GCGGCCGATG

41201 TCAGTGATCC GGACGCCGTG GCGGAGCTCG TCGGCAAGAC CGATCCGGCG

41251 CACCCGCTGA CCGGTGTGAT CCACGCGGCC GGTGTGCTGG ACGACGCCGT

41301 GGTCACCGCA CAGACCCCGG AGAGCCTCGC GCGGGTGTGG GCGGCGAAGG

41351 CGACGGCCGC ACACCTGCTG CACGAGGCGA CCCGGGAGGC GCGCCTCGGT

41401 CTCTTCCTGG TGTTCTCCTC GGCGGCGGCG ACACTCGGCA GTCCGGGACA

41451 GGCCAACTAC GCGGCGGCCA ACGCCTATTG CGACGCCCTC GTCCGGCAAC

41501 GGCGTGCCGA GGGCCTGGCC GGTCTCTCGA TCGGCTGGGG TCTGTGGCAG

41551 ACGGCGAGCG GCATGACCGG ACACCTCGGC GAGACGGACC TGGCACGCAT

41601 GAAGCGCACC GGGTTCACCC CGCTGACCAC CGAAGGTGGC TTGGCCCTCC

41651 TCGACGCCGC CCGCGCCCAC GGCCGCCCGC ACGTGGTCGC GGTGGACCTC

41701 GACGCGCGCG CCGTCGCCGC GCAGCCCGCC CCGTCCCGGC CCGCGCTCCT

41751 GCGCGCCCTG GCCGCGGGTG CGACCCCGGG GGCCCGCACC GCCCGGCGCA

41801 CCGCGGCCGC GGGCAGCGTC GCCCCGGCGG GCGGTCTCGC CGACCGGCTC

41851 GCCGGCCTGC CGCATCCCGA ACGGCGCCGG CTGCTGCTCG ACCTCGTACG

41901 TGGCAACGTC GCCGGCGTCC TCGGGCACAG CGACCACGAC GCCGTCCGCC

41951 CGGACACGTC GTTCAAGGAG CTCGGCTTCG ACTCCCTGAC CGCCGTGGAA

42001 CTGCGCAACC GGCTGGCCGC CGCCACCGGC CTGAAGCTGC CCGCGGCGCT

42051 CGTCTTCGAC TACCCCGAGT CGGCCACCCT CGTCGACCAC CTCCTGGAGC

42101 GTCTGTCGCC CGACGGCGCG CCGCCGCCCG TCAAGGACGC CGCGGACCCC

42151 GTTCTCAACG ACCTCGGCAG GATCGAGTCC TCCCTGGACG CGCTCGCCCT

42201 CGACGCGGAC GCGCGCAGCC GGGTCACCAG GCGTCTGAAC ACCCTGCTGT

42251 CGAAGCTGAA CGGAGCCGCC ACCGCCGGCT CCCCGGCGGA CGTCACGGAC

42301 CTGGACGCGC TGGACGCGCT GGACGACGTG TCCGACGACG AGATGTTCGA

42351 GTTCATCGAC CGAGAGCTGT GACCCCCCTG CCCGCCCCGT CCCCCTTCCC

42401 CGCCCCCACG TTCCCCGTGC CCTTCGCTGA TGGAGAAGTG ACGTTCGATG

42451 TCGAGTGCTG AAGAGTCGAG TCCTGATGTG TCCGGCACGG GTGTGTCCGG

42501 TACGGGAGAG TCCGCTACGG GTACGTCGAG TACGGAAGCC AAGCTTCGGC

42551 AGTATCTGAA GCGGGTCACG GTGGACCTCG GCCAGGCCCG CCGGCGGCTG

42601 CGCGAGGTGG AGGAGCGGGC CCAGGAGCCG ATCGCCATCG TCTCCATGGC

42651 GTGCCGCTTC CCCGGCGACA CCCGCACGCC CGAGGCCCTG TGGGACCTGG

42701 TCGCCGAGGG CGGCGACGCC ATCGACGACT TCCCCACCAA TCGCGGCTGG

42751 GACCTGGAGA GCCTCTACCA CCCCGACCCC GACCACCCCG GCACCAGCTA

42801 CGTCCGACGC GGCGGGTTCC TGTACGACGC CCCCGCCTTC GACGCGTCGT

42851 TCTTCGGGAT CAGCCCGCGC GAAGCCCTGG CCATGGACCC GCAGCAGCGG

42901 GTGCTCATGG AGACGGCCTG GCAGCTCCTG GAGCGGGCCG GCATCGACCC

42951 GGCCTCGCTG AAGCTGAGCG CCACCGGCGT CTACATCGGC GCGGGCGTGC

43001 TCGGGTTCGG CGGCGCGCAG CCCGACAAGA CGGTAGAGGG CCACCTCCTG

43051 ACCGGCAGCG CGCTGAGTGT CCTGTCCGGC CGCATCTCCT TCACGCTCGG

43101 CCTCGAGGGC CCGTCGGTCA GTGTCGACAC GGCGTGCTCC TCCTCGCTGG

43151 TCTCCATGCA CCTGGCGGCC CAGGCGCTGC GGCAGGGGGA GTGCGATCTC

43201 GCGCTGGCCG GCGGTGTCAC CGTGATGTCG ACGCCCGGCG CGTTCACCGA

43251 GTTCTCCCGC CAGGGCGCGC TGTCTCCGGA CGGCCGCTCG AAGGCTTTCG

43301 CGGCCTCGGC CGACGGCACC GGTTTCTCGG AGGGCGCGGG ACTGCTCCTC

43351 CTGGAGCGGC TCTCCGACGC GCGCCGCAAC GGCCACAAGG TGCTCGCGGT

43401 GATCCGCGGC TCGGCCGTCA ACCAGGACGG CGCGAGCAAC GGTCTCACCG

43451 CCCCCAACGG CCCCTCCCAG GAACGCGTGA TCCGCGCCGC CCTCGCCAAC

43501 GCGGGCCTGG GCGCCGCCGA GGTCGACGCG GTCGAGGCAC ACGGCACCGG

43551 CACGAAGCTC GGCGACCCCA TCGAGGCCGG TGCGCTGCTC GCCACCTACG

43601 GCCGCGACAG GGACGAGGAC CGGCCGCTGT GGCTGGGCTC GGTCAAGTCG

43651 AACATCGGTC ACCCGCAGGG CGCAGCAGGC GTCGCGGGCG TCATCAAGAT

43701 GGTGATGGCG CTGCAGCGCG AACTGCTCCC CGCCACCCTG TACGTCGACG

43751 AGCCCACCCC GCACGTCGAC TGGTCCTCGG GCTCCGTCAG GCTCCTCACC

43801 GAACCGGTCC CGTGGACCCG CGGCGAGCGC CCGCGCCGCG CGGGCGTGTC

43851 CGCCTTCGGC ATGTCCGGGA CGAACGCCCA CGTGATCCTG GAGGAGGCAC

43901 CGCCCGAGGA GGCAGCGGCC GCGGAGACAC CGGCGGAAGG GACAGGCGCA

43951 GTCGTCCCGT GGGTCGTCTC CGGCCGGGGC GAGGAAGCGC TGCGGGCCCA

44001 GGCCGCACAG CTCGCCGAGC ACGTGCGCGA CGACGACCAG CGGCCGGCGT

44051 CACCGCTGGA GGTGGGGTGG TCGCTCGCCA CGACACGGTC GGTGTTCGAG

44101 AACCGGGCCG TCGTCGTCGG GGACGACCGC GACGCGCTCC TCGACGGCCT

```
44151  CCGGTCGCTG GCGGCAGGTG AGGCGTCGCC GGACGTGGTG TCCGGGGCGG
44201  TCGGCCCCAC GGGGCCCGGG CCGGTCATGG TGTTCCCCGG CCAGGGCGGC
44251  CAGTGGGTGG GCATGGGGGC CCGGCTCCTC GACGAGTCCC CGGTGTTCGC
44301  GGCCCGGATC GCCGAGTGCG AGCAGGCCCT GTCGGCGTAC GTGGACTGGT
44351  CCCTGACCGA CGTGCTGCGC GGGGACGGGT CGGAGCTGGC CCGGATCGAC
44401  GTCGTCCAGC CCGTGCTGTG GGCCGTCATG GTCGCGCTCG CCGCCGTCTG
44451  GGCGGACCAG GGAATCGAAC CCGCCGCCGT CGTCGGCCAC TCGCAGGGCG
44501  AGATAGCCGC GGCGTGCGTC GTGGGCGCCA TCTCCCTGGA CGAGGCGGCC
44551  CGCATCGTCG CCGTACGCAG TGTGCTGCTG CGGCAGCTGT CCGGACGCGG
44601  CGGCATGGCG TCCCTGGGGA TGGGCCAGGA GCAGGCCGCC GACCTGATCG
44651  ACGGACACCC GGGTGTGGTC GTCGCGGCCG TCAACGGGCC GTCGTCCACC
44701  GTCATCTCGG GCCCGCCCGA GGGCATCGCC GCCGTCGTCG CCGACGCCCA
44751  GGAGCGGGGC CTTCGCGCCA GGGCCGTCGC CTCCGACGTC GCGGGCCACG
44801  GCCCGCAGCT GGACGCGATC CTGGACCAGC TCACGGAGGG CCTGGCCGGC
44851  ATCCGGCCCG CCGCGACCGA CGTCGCGTTC TACTCCACCG TCACCGCCGG
44901  GCACCTCACC GACACCACCG AACTCGACAC CGCGTACTGG GTGCGGAACG
44951  TGCGCCGGAC GGTGCGTTTC GCCGACACGA TCGACGCGCT GCTCGCGGAC
45001  GGGTACCGCC TGTTCATCGA GGTGAGCCCC CACCCCGTCC TCAACCTCGC
45051  GCTGGAAGGC CTCATCGAAC GGGCGGCCGT GCCCGCCACG GTCGTGCCCA
45101  CCCTGCGCCG CGACCACGGC GACACCACCC AGCTCGCCCG CGCCGCGGCC
45151  CACGCCTTCG CCGCCGGCGC GGACGTCGAC TGGCGGCGCT GGTTCCCGGC
45201  CGACCCCGCC CCCCGTACCG TCGACCTGCC CACCTACGCC TTCCAGCGCC
45251  AGGACTTCTG GCCGGCCCCC GCCGGCGGGC GGTCCGGCGA CCCTGCCGGG
45301  CTCGGCCTCG CCGCCTCCGG ACACCCGCTC CTGGGCGCCT CCGTGGGCCT
45351  CGCGAGCGGG GACGTACACC TGCTGAGCGG GCGGGTGTCC CGGCAGTCCG
45401  CCGCGTGGCT GGACGACCAC GTCGTGGCGG GCCAGGCCCT GGTGCCCGGC
```

```
45451  GCGGCGCAGG TGGAGTGGGT GCTGCGGGCC GGCGACGACG CGGGCTGCTC
45501  CGCCCTGGAG GAGCTGACGC TCCAGACGCC GCTCGTGCTG CCCGACACCG
45551  GCGGCCTGCG GATCCAGGTC GTCGTCGAAG CGGCCGACGC ACACGGCCGG
45601  CGCGACGTCC GGCTGTTCTC CCGCCCCGAT GACGACGACG CCTTCGCGTC
45651  GACGCACCCC TGGACCTGCC ACGCCACGGG CGTGCTCGCC CCCGCCCCGA
45701  CGGACGGCAC CAACGGAACG CGGGACGCCG CCGACACCCT GGACGGCGCA
45751  TGGCCCCCGG CCGACGCCGA ACCCGTCCCC GCCGACGACC TCTACGCGCA
45801  GGCCGACCGC ACCGGATACG GCTACGGCCC CGCCTTCCGG GGCGTACGGG
45851  CGCTGTGGCG CCACGGCAAG GACGTCCTGG CCGAGGTGAC GCTGCCCAAG
45901  GAGGCCGGCG ACCCGGACGG CTTCGGTATC CACCCGGCCC TCCTCGACGC
45951  CGTCCTGCAA CCCGCCGCAC TGCTGCTGCC CCCGACCGAC GCCGAACAGG
46001  TCTGGCTGCC GTTCGCCTGG AACGACGTGG CGCTGCACGC CGTACGGGCC
46051  ACCACGGTCC GGGTGCCGCT CACCCCGCTC GGCGAGCGGA TCGACCAGGG
46101  GCTGCGCATC ACCGTGGCCG ACGCCGTGGG CGCGCCCGTG CTCACCGTCC
46151  GCGACCTGCG CTCGCGCCCG ACCGACACAG GCCGCCTCGC CGCGGCCGCG
46201  ACCCGCGACC GGCACGGGCT GTTCGACCTG GAGTGGATCG CGCCGGAGAA
46251  CGCGGCGGAG AACGCGGCGG GTCCGGCCCG GGACGCGTCC GAAGGGTGGG
46301  TGACACTCGG CGAGGACGCC GCGAGCCTCG CGGACCTGCT GGCGTCCGTC
46351  GAGGCGGGCG CTCCGGCGCC GCAGCTCGTG GCCGCCCCCG TCGAACCCGA
46401  CCGGACCGAC GACGGCCTGG CACTCGCCAC CCACGTCCTC GACCTCGTAC
46451  AGACCTGGCT CGCCTCGCCC CTGCACGACT CCCGCCTGGT CCTGGTGACG
46501  CGAGGGGCAG TGACGGATGC GGATGTGGAT GTGGCTGCCG CGGCCGTTTG
46551  GGGTCTGGTA CGCAGCGCCC AGTCGGAGCA CCCCGGCCGC TTCACGCTGA
46601  TCGACCTCGG CCCCGACGAC ACGCTTGCCG CAGCCATGCA GGCGGCGCAC
46651  CTGGAAGAGC CGCAACTGGC GGTGCACGGC GGCGAGATAC GAGTGCCGCG
46701  ACTGGTCCGC GCCACGACCG ACCCGACCGC CCCGAACGGG ACACCGGAGG
```

```
46751  CCGACCGGAC GGCGGACCCG TCCGAAGGAC TCCACCGGAA CGGTACGGTT
46801  CTCATCACCG GCGGCACCGG CGTACTCGGC CGACTGGTGG CCGAACACCT
46851  GGTCACGGAG TGGGGCGTAC GCCACCTGCT GCTCGCGAGC CGACGCGGCG
46901  ACCAGGCGCC GGGTAGCGCC GAACTCCGCG CCCGCCTGAG CGAATTGGGA
46951  GCATCGGTCG AGATCGCCCC GGCCGATGTC GGCGACGCGG AAGCGGTCGC
47001  CGCACTGATC GCGTCGGTCG ACCCGGCGCA CCCGCTCACC GGTGTGATCC
47051  ACGCGGCCGG TGTCCTGGAC GACGCCGTGA TCACCGCCCA GACCCCCGAG
47101  AGCCTCGCGC GGGTGTGGGC GACGAAGGCG ACGGCGGCCC GCCATCTGCA
47151  CGAGGCGACA CGGGAGACAC CCCTCGACTT CTTCGTGGTG TTCTCCTCGG
47201  CGGCCGCCTC GCTCGGCAGC CCCGGCCAGG CCAACTACGC GGCGGCCAAC
47251  GCCTATTGCG ACGCCCTCGT CCAGCACCGC CGCGCCCAAG GGCTCGCGGG
47301  CCTCTCGATC GCCTGGGGCC TGTGGCAGGC GACCAGCGGC ATGACCGGGC
47351  AGCTGAGCGA GACCGACCTG GCGCGCATGA AGCGCACCGG GTTCGCCGCG
47401  CTGACCGACG AGGGCGGCCT GGCCCTGCTC GACGCCGCCC GTGCCCACGA
47451  CCGGGCCTAC GTGGTCGCGG CCGACCTCGA CCCGCGCGCC GTGACCGATG
47501  GCCTGTCCCC GCTCCTGCGC GCCCTCACGG CGCCCGCCAC GCGGCGGCGC
47551  GTGGCCTCCG AAGGCCTCGC CGACGGGGCG CTCGCGACCC GCCTGGCCGG
47601  CCTCGACGCG GACGGCCGCC TAAGGCTCCT CACCGATGTC GTACGCGAGT
47651  ACGTCGCGGC CGTCCTCGGC CATGGTTCCG CCGCCCGGGT GGGCGTCGAC
47701  ATCGCCTTCA AGGACCTGGG TTTCGACTCG CTGACCGCGG TGGAGCTGCG
47751  CAACCGGCTG TCGGCCGCCT GTGACGTGCG GCTGCCCGCC ACACTGATCT
47801  TCGACCACCC CACCCCGCAG GCTCTCGCCA CCCACCTGGT GGACCGCTTG
47851  GCGGGCAGCA CCTCCGCGAC CACGACGGTG AATGCGACGG CGCCGGCAGC
47901  CGCCCACGTC GCCGCAGGGG CCGACGTCGA CGCAGACACC GACGACCCGG
47951  TCGCCATCGT CGCCATGACG TGCCGGTTCC CGGGCGGCGT CGCGTCCCCG
48001  GACGACCTGT GGGACCTGCT CGACGCACGC AAGGACGCGA TGGGCGCCTT
```

```
48051  CCCCACCGAC CGCGGCTGGG ACCTGGAACG CCTCTTCCAC CCCGACCCGG
48101  ACCACCCCGG CACCAGCTAC ACCGACCAGG GCGGATTTCT TCCCGACGCG
48151  GGTGATTTCG ATGCGGCGTT CTTCGGGATC AATCCGCGGG AGGCGCTGGC
48201  GATGGATCCG CAGCAGCGGT TGTTGCTGGA GGCGTCGTGG GAGGTGTTGG
48251  AGCGTGCGGG TATCGATCCG ACGACGCTCA AGGGCACCCC GACCGGCACC
48301  TACGTGGGCC TCATGTACCA CGACTACGCC AAGTCCTTCC CCACGGCCGA
48351  CGCCCAGTTG GAGGGCTACT CCTACTTGGC GAGCACCGGC AGCATGGTCT
48401  CCGGCCGCGT CGCCTACACC CTGGGCCTTG AAGGTCCGGC GGTGACGGTC
48451  GACACCGCGT GCTCCTCCTC CCTGGTCTCC ATCCACCTGG CGACGCAGGC
48501  ACTCCGGCAC GGCGAGTGCG ACCTCGCCCT GGCAGGCGGT GTGACCGTCA
48551  TGGCCGACCC GGACATGTTC GCGGGCTTCT CGCGCCAGCG CGGCCTCTCA
48601  CCTGACGGCC GCTGCAAGGC CTACGCCGCC GCGGCCGACG GAGTCGGATT
48651  CTCCGAGGGA GTGGGCGTAT TGCTCCTTGA GCGGTTGTCG GATGCGCGGC
48701  GTCATGGGCG TCGGGTGTTG GGTGTGGTGC GGGGTTCGGC GGTGAATCAG
48751  GACGGTGCGA GTAATGGGTT GACGGCGCCG AATGGTCCGT CGCAGGAGCG
48801  GGTGATTCGT CAGGCGTTGG CCAGTGGTGG GTTGTCGTCG GTGGATGTTG
48851  ATGTGGTGGA GGGGCATGGG ACGGGGACCA CGTTGGGTGA TCCGATCGAG
48901  GCGCAGGCTC TGCTGGCCAC ATATGGGCAG GGGCGTCCGG AGGACCGTCC
48951  GTTGTGGTTG GGGTCGGTGA AGTCGAACAT TGGTCATACG CAGGCGGCTG
49001  CGGGTGTTGC GGGTGTCATC AAGATGGTGA TGGCGATGCG GCATGGTGTG
49051  GTGCCGGCGA GTTTGCATGT GGATGTGCCG TCGCCGCATG TGGAGTGGGA
49101  TTCGGGTGCG GTGCGGTTGG CGGTTGAGTC GGTGCCATGG CCGCAGGTGG
49151  AGGGTCGTCC GCGTCGGGCG GGTGTGTCGT CGTTCGGCGC TTCGGGGACG
49201  AATGCGCACG TGATCGTGGA GTCTGTTCCC GATGGGCTGG AGGAGGACTC
49251  GGTATCGGTC GGCGGTGAGG CTCTTGAGAC GGAGACTGAC GGGCGCTTGG
49301  TGCCGTGGGT GGTGTCGGCC CGCAGCCCGC AGGCCCTGCG CGACCAGGCA
```

```
49351  CTACGCCTGC GTGACTTTGC CAGTGACGCG TCGTTCCGCG CGCCGCTCGC
49401  CGACGTGGGC TGGTCGCTGC TGAAGACGCG TGCGCTGCAT GAGCATCGCG
49451  CCGTTGTGGT GGGCGCGGAG CGGGCAGAGC TGATCGCCGC TCTGGAGGCG
49501  CTGGCGACGG GTGAGCCGCA TGCGGCGCTG GTCGGCCCGG CTTGCTCGCA
49551  GGCTCGGGTG GGTGGCGATG ACGTGGTGTG GCTGTTCAGT GGTCAGGGCA
49601  GTCAGTTGGT CGGTATGGGT GCTGGTTTGT ATGAGCGGTT CCCGGTGTTT
49651  GCGGCTGCGT TTGATGAGGT GTGCGGCCTG TTGGAGGGGC CGTTGGGCGT
49701  GGAGGCGGGT GGGTTGCGGG AGGTGGTGTT CCGTGGCCCG CGGGAGCGGT
49751  TGGATCACAC GGTGTGGGCG CAGGCGGGGT TGTTTGCGCT GCAGGTGGGG
49801  TTGGCCCGGT TGTGGGAGTC GGTCGGGGTG CGGCCGGATG TGGTGCTCGG
49851  GCATTCGATC GGTGAGATCG CGGCCGCGCA TGTGGCGGGG GTTTTTGATC
49901  TGGCGGATGC GTGTCGGGTG GTGGGTGCGC GGGCGCGTTT GATGGGTGGG
49951  CTGCCTGAGG GTGGGGCGAT GTGCGCGGTG CAGGCCACGC CCGCCGAGCT
50001  GGCCGCCGAC GTGGACGGAT CGGCTGTAAG TGTGGCGGCA GTCAACACCC
50051  CCGACTCCAC GGTGATTTCG GGCCCGTCGG ACGAGGTGGA CCGGATTGCT
50101  GGGGTGTGGC GGGAGCGTGG GCGCAAGACG AAGGCGCTGA GCGTCAGTCA
50151  TGCCTTCCAT TCGGCGTTGA TGGAGCCGAT GCTCGCGGAG TTCACCGAAG
50201  CGATACGAGG GGTCAAGTTC AGGCAGCCGT CGATCCCGCT CATGAGCAAT
50251  GTCTCCGGAG AGCGGGCCGG CGAGGAGATC ACGGATCCGG AGTACTGGGC
50301  GAGGCATGTA CGTAATGCGG TGCTCTTCCA GCCCGCCATC GCCCAAGTAG
50351  CGGATTCAGC GGGCGTGTTT GTGGAGCTCG GCCCCGCGCC TGTGCTGACC
50401  ACGGCCGCCC AGCACACCCT GGACGAGTCG GACAGCCAGG AGTCGGTGCT
50451  GGTCGCGTCT CTCGCCGGTG AGCGTCCTGA GGAGTCGGCG TTTGTGGAGG
50501  CGATGGCTCG TCTGCATACC GCTGGTGTTG CTGTGGACTG GTCGGTGTTG
50551  TTCGCGGGTG ATCGTGTGCC TGGGCTGGTG GAGTTGCCGA CGTATGCGTT
50601  CCAGCGGGAG CGGTTCTGGT TGAGTGGCCG TTCTGGGGGT GGGGATGCGG
```

50651 CGACTTTGGG GTTGGTGGCG GCGGGGCATC CGTTGTTGGG TGCGGCGGTG

50701 GAGTTCGCGG ACCGGGGTGG GTGTCTGCTG ACCGGTCGTC TGTCGCGGTC

50751 TGGGGTGTCG TGGCTTGCTG ATCATGTGGT GGCGGGTGCG GTTTTGGTGC

50801 CGGGTGCTGC GTTGGTGGAG TGGGCGTTGC GGGCCGGTGA TGAGGTCGGT

50851 TGTGTGACGG TGGAGGAGTT GATGTTGCAG GCGCCTTTGG TGGTGCCTGA

50901 GGCGTCGGGT CTGCGGGTTC AGGTGGTGGT TGAGGAGGCG GGTGAGGACG

50951 GGCGGCGCGG TGTTCAGATC TACAGCCGGC CCGACGCGGA CGCCGTGGGC

51001 GGCGATGACT CGTGGATCTG CCACGCGACC GGCGTACTGT CACCCGAAAG

51051 CGCTCGTCTG GACACGGAGT TGGGTGGCGT CTGGCCACCG GCCGGTGCCG

51101 AACCGCTGGA TGTCGACGGC TTCTACGCGC AGGCCGGTGA GGCCGGGTAC

51151 GGATACGGTC CGGCGTTCCG GGGGCTGCGT GCCGTGTGGC GGCACGGCCA

51201 GGACCTGCTG GCCGAGGTCG TCCTGCCCGA AGCCGCCGGT GCCCATGACG

51251 GCTACGGGAT CCACCCCGCC CTCCTCGACG CCACCCTCCA TCCGCTGCTC

51301 GCCGCCCGCT TCATGGACGG TTCCGAGGAC GATCAGCTCT ACGTACCGTT

51351 CGGGTGGGCC GGAGTGTCTC TGCGGGCGGT GGGAGCCACG ACTGTGCGCG

51401 TGCGCCTCCG TCCGGTCGGG GAGAGCGTCG ACCAAGGGCT GAGCGTGACG

51451 GTCACCGATG CGACCGGCGG TCCCGTTCTG AGCGTCGACT CCCTCCAGAC

51501 CCGCCCCGTG AAGCCGAGCC AATTGGCTGC GGCCCAACAG CCGGACGTAC

51551 GCGGTCTGTT CACTGTGGAG TGGACGCCGC TGCCGCAGAC GGATGCCGAC

51601 GGGGAGGCCG ACTGGGTTGT GCTCTCGGAC GGTGTTGGCC GTCTGGCTGA

51651 TGTGGTGTCG GCGGCGGGTG GTGAAGCGCC GTGGGCAGTG GTCGCTCCTG

51701 TCGATGCGTC TGTGGGCGAC GGCCGTGAGG GTCTTGACGG TCGGCTGGTC

51751 GTGGAGCGGG TGCTGTCACT CGTACAGGAG TTCCTGGCCC TGCCGGAGCT

51801 GGCCGAGTCC CGTCTCCTCG TGGTGACGCG CGGTGCGGTG GCCACCGGCG

51851 TCGACGGTGA CGGTGACGTG GACGCGTCCG CCGCAGCTGT ATGGGGCCTG

51901 GTCCGCAGTG CTCAGTCCGA GAATCCGGGC CGCTTCATCC TGCTCGACGT

```
51951  GGACGGCGAC GGCGACGACC AGGGCCCGGA CCTGAACGGC CGGCATCTGC
52001  CCCACGCCAC CCTGCGTCAC GCCGCCGAGG AACTCGACGA GCCCCAACTC
52051  GCCCTGCGGG AAGGGACGCT CTACGTCCCC CGACTGACCC AGGCGCGCCA
52101  GTCCGCCGAA CTCGTCGTGC CGCCCGGTGA ACCGGCGTGG CGCCTGCGGA
52151  TGGTGCACGA CGGCTCGCTG GACGCCCTGG CGGCAGTGGC CTGCCCGGAG
52201  GCCCTGGAGC CCTTGGCGCC GGGGCAGGTG CGTATCGCCG TACACGCCGC
52251  GGGCATCAAC TTCCGTGACG TACTGGTGGC CTTGGGTATG GTCCCCGCGT
52301  ACGGGGCCAT GGGTGGCGAA GGTGCCGGTG TCGTGACGGA GGTCGGTCCC
52351  GAGGTCACCC ATGTCTCGGT GGGCGACCGC GTGATGGGCG TGTTCGAGGG
52401  CGCGTTCGGC CCTGTGGTGA TCGCCGAGGC GCGGATGGTC ACACCTGTCC
52451  CGCAGGGCTG GGACATGCGG GAGGCGGCCG GTATTCCGGC GGCCTTCCTG
52501  ACGGCTTGGT ACGGGTTGGT GGAGCTGGCC GGTCTGAAGG CGGGCGAGCG
52551  GGTGCTGGTC CATGCCGCGA CGGGTGGTGT GGGGATGGCG GCGGTGCAGA
52601  TCGCCCGGCA TGTGGGTGCC GAGGTGTTCG CCACCGCGAG TCCGGGCAAG
52651  CACGCCGTGC TGGAGGAGAT GGGCATCGAC GCCGCCCACC GCGCCTCCTC
52701  CCGGGACCTC GCCTTCGAGG GCACGTTCAG GGAAGCAACG GGCGGCCGCG
52751  GCATGGACGT CGTGCTCAAC AGCCTTGCCG GCGAGTTCAT CGACGCCTCT
52801  CTGCGGTTGC TCGGCGACGG CGGCCGGTTC CTGGAGATGG GCAAGACCGA
52851  TGTGCGGGCC GCCGAAGAGG TGGCTGCGGA GCACGCGGAC GTCTCGTACA
52901  CGGCGTACGA CCTCGTCGGT GATGCCGGAC CCGACCGCAT CAGCAACATG
52951  CTGGACAAGC TCGTCGAATT GTTCGCCTCA GAACGGCTTA AGCCGCTGCC
53001  GGTACGTTCC TGGCCGCTGG ACAAGGCGCA GGAGGCGTTC CGGTTCATGA
53051  GTCAGGCGAA GCACACCGGC AAGCTGGTGC TTGAGATCCC GCCTGCCCTC
53101  GACCCCGAGG GCACGGTTCT GGTCACCGGG GGCACCGGTG CGCTGGGGCA
53151  GGTCGTGGCC GAGCATCTGG TCCGGGAGTG GGGCGTACGG CACCTGCTGC
53201  TGGCCAGCCG TCGCGGTCCG GAGGCGCCGG GCAGCGACGA ACTGGCCTCG
```

53251 AAGCTCACCG GGTTGGGTGC CGAGGTCACC ATTGTCGCGG CCGATGTCAG

53301 CGACCCGGCC TCGGTGGTGG AGCTGGTCGG CAAGACGGAT CCCTCGCATC

53351 CGTTGACGGG TGTCGTGCAC GCGGCGGGCG TGTTGGAGGA CGGTGTCGTG

53401 ACCGCTCAGA CGCCTGAGGG GCTGGCGCGG GTGTGGGCGG CCAAGGCTGC

53451 TGCGGCGGCG AATCTCCATG AGGCGACCCG GGAGATGCGT CTCGGCCTGT

53501 TCGTGGTGTT CTCCTCGGCG GCCGCCACGC TCGGCAGTCC GGGCCAGGCC

53551 AACTACGCGG CCGCCAATGC CTATTGCGAC GCGCTGATGC AGCACCGACG

53601 GGCGGTGGGC CAGGTCGGCC TGTCGGTCGG CTGGGGTCTC TGGGAGGCGC

53651 CGGACGCCAA GCCGGGTGTT GCCGCCGACG CCAAGGCGAG TGCTGCCACC

53701 GTCGGCAAGG CGAGTGCTCT ATCCGACGGC ACGAACGGCA GCGCTCCCCA

53751 GGACACGACC GGCACCGCCC CCCAGGGCAT GACCGGCGGA CTCACCGACA

53801 CCGACGTAGC CCGCATGGCA CGTATCGGCG TCAAGGGCAT GAGCAACGCC

53851 CACGGTCTCG CCCTGTTCGA CGCCGCGCAC CGCCACGGCC GCCCCCACCT

53901 GGTCGGCTTC AACCTCGACC TGCGCACCCT GGCCACGCAC CCCCTGCACA

53951 CCCGGCCCGC CCTTCTGCGC GGCCTGGCCA CCCCCACCGC CGGCGGGGCG

54001 AGCAGGCCGA CCGCGACGGC GGGCGGACAG CCCGCCGACC TGGCGGGCCG

54051 GCTGGCCGCG CTGTCGCCGT CGGACCGGCA CCACACGCTG GTCCGGCTCA

54101 TCAGGGAACA GGCCGCCACC GTGCTCGGGC ACCACCCGGA CAGTCTCACC

54151 ACGGGCAGCA CCTTCAAGGA ACTCGGATTC GACTCCCTGA CCGCGGTCGA

54201 ACTGCGCAAC AGGCTGTCCG CCGCCACCGG TCTCCGGCTC CCCGCCGGCC

54251 TGGTCTTCGA CCACCCGGAC GCCGACATCC TGGCCGAACA CCTCGGCGCG

54301 CAACTCGCCC CCGACGGGGA CACCCCCGCC GGTGCGGAAG CCACCGACCC

54351 GGTCCTCCGC GACCTGGCGA AACTCGAGAA CGCCCTCTCC TCCACCCTCG

54401 TCGAGCACCT CGACGCCGAC GCGGTCACGG CCCGACTGGA AGCACTCCTG

54451 TCGAACTGGA AGGCGGCGAG CGCGGCGCCC GGCTCGGGCA GCACGAAGGA

54501 GCAGCTCCAG GTTGCCACGA CCGACCAGGT CCTCGACTTC ATCGACAAAG

54551 AACTGGGTGT GTGAAACGAC CGTGCACGGC GCGACAACCA CGCTGAAGGC

54601 TGGGTGAACT CTCATGGCGA GTGAAGAGGA ACTGGTCGAC TACCTCAAGC

54651 GGGTCGCCGC CGAACTGCAC GACACCCGGC AGCGCCTGCG CGAGGTCGAG

54701 GACCGGCGGC AGGAGCCGGT GGCCGTCGTC GGCATGGCCT GCCGTTTCCC

54751 CGGCGGCATC GAGACGCCCG AGGGACTGTG GGAGCTGGTC GCGGCCGGCG

54801 ACGACGCCAT TGAGCCCTTC CCCACCGACC GGGGCTGGGA CCTGGAAGGC

54851 ATCTACCACC CGGACCCCGA CCACCCGGGT ACCTGCTACG TCCGGGAGGG

54901 CGGGTTCCTA GCCGCCCCTG ACCGGTTCGA CTCCGACTTC TTCGGCTTCA

54951 GCCCGCGCGA GGCCCTGGCC AGCAGCCCGC AACTGCGACT GCTCCTGGAG

55001 ACGTCCTGGG AGGCCCTCGA ACGGGCGGGC ATCAACCCCG CCTCGCTCAA

55051 GGGCAGCCCC ACCGGCGTCT ACGTCGGCGC CGCGACCACC GGCAACCAGA

55101 CGCAGGGCGA CCCCGGCGGC AAGGCGACCG AGGGTTACGC GGGCACCGCG

55151 CCCAGCGTCC TCTCGGGCCG CCTCTCGTTC ACGCTCGGCC TGGAGGGCCC

55201 GGCGGTGACC GTCGAGACAG CGTGCTCCTC CTCGCTGGTG GCGATGCACC

55251 TGGCGGCCAA CGCCCTGCGC CAGGGCGAGT GCGACCTCGC CCTCGCGGGC

55301 GGCGTCACCG TCATGTCCAC CCCCGAGGTG TTCACAGGCT TCTCGCGTCA

55351 GCGGGGACTG GCCCCCGACG GCCGCTGCAA GCCGTTCGCC GCCGCGGCCG

55401 ACGGCACGGG CTGGGGCGAG GGCGCGGGCC TGATCCTCCT GGAGCGCCTC

55451 TCCGACGCCC GCAGGAAGGG CCACAAGGTC CTCGCGGTGA TCCGGGGCTC

55501 GGCGATCAAC CAGGACGGCG CGAGCAACGG CTTCACCGCG CCCAACGGCC

55551 CCTCGCAGCG CCGCGTCATC CGCCAGGCAC TCTCCAGCGC CCACCTCTCC

55601 ACGTCGGAGA TCGACGTCGT CGAGGCGCAC GGCACCGGCA CCAGGCTCGG

55651 CGACCCCATC GAGGCCGAGG CGCTCATCGC CACCTACGGC AAGGAGCGCG

55701 AGGACGACCG TCCCCTGTGG CTCGGCTCGG TCAAGTCCAA CATCGGCCAC

55751 ACGCAGGCCG CCGCGGGCGT CGCCGGAGTC ATCAAGATGG TGATGGCGCT

55801 ACAGCGCGAA CTGCTTCCCG CCACCCTGAA CGTCGACGAG CCGACCCCGC

55851 ACGTCCAGTG GGAGGGCGGC GGCGTACGCC TCCTGACCGA ACCGGTCCCG

55901 TGGTCGCGCG GCGAACGCCC GCGCCGCGCC GGAATCTCCT CCTTCGGCAT

55951 ATCGGGCACG AACGCGCACG TGGTCCTGGA GGAGGCGCCG CCGGAGGAGG

56001 ACGTGCCGGG CCCCGTGGCT GCGGAGCCGG AAGGGGTGGT GCCGTGGGTG

56051 GTCTCCGCGC GGACCGAGGA GGCGTTGAGC GAACAGGCGC GGCGCCTGGG

56101 CGAGTTCGTG GCCGACACGG ACCCGTCGAC CGCTGACGTC GGGTGGTCAC

56151 TGACCACGAG CAGGGCGATC CTTGAACACC GCGCTGTGGT GGTGGGGCGT

56201 GATCGGGATG CGCTGACGGC CGGCCTGGCG GCGTTGGCCG CGGGTGAGGA

56251 GTCGGCGGAT GTGGTGGCTG GGGTGGCCGG TGATGTGGGT CCTGGGCCGG

56301 TGTTGGTGTT TCCGGGGCAG GGGTCGCAGT GGGTGGGCAT GGGCGCCCAG

56351 CTCCTTGACG AGTCGCCCGT CTTCGCGGCG CGGATCGCGG AGTGTGAGCA

56401 GGCGCTGTCG GCGTACGTGG ACTGGTCGCT GAGTGCGGTG TTGCGCGGGG

56451 ATGGGAGTGA ACTGTCCCGG GTCGAGGTCG TGCAGCCGGT GTTGTGGGCG

56501 GTGATGGTCT CGCTGGCTGC CGTCTGGGCG GATTACGGGG TCACCCCGGC

56551 CGCTGTGATC GGGCACTCGC AGGGCGAGAT GGCCGCCGCG TGCGTGGCGG

56601 GGGCGCTGTC TTTGGAGGAT GCGGCGCGCG TCGTGGCCGT ACGCAGTGAC

56651 GCGCTTCGTC AGCTGATGGG GCAGGGCGAC ATGGCGTCGT TGGGCGCCAG

56701 CTCGGAGCAG GCGGCTGAGC TCATCGGTGA TCGGCCGGGC GTATGCATCG

56751 CAGCGGTCAA CGGGCCGTCC TCGACAGTCA TTTCAGGACC GCCGGAGCAT

56801 GTGGCAGCCG TGGTCGCGGA TGCGGAGGAA CGTGGTCTGC GCGCCCGTGT

56851 CATCGATGTC GGCTATGCCT CGCACGGTCC CCAGATCGAT CAGCTCCACG

56901 ACCTCCTCAC CGACCGGCTC GCCGACATCC GGCCCGCGAC CACGGACGTG

56951 GCCTTCTATT CGACGGTCAC CGCCGAGCGC CTGACGGACA CCACGGCCCT

57001 GGATACGGAT TACTGGGTTA CCAACCTCCG CCAGCCGGTC CGTTTCGCCG

57051 ACACCATCGA TGCGCTTCTC GCGGACGGCT ATCGCCTGTT CATCGAGGCC

57101 AGCGCGCACC CGGTGCTGGG TCTGGGCATG GAGGAGACCA TCGAGCAGGC

57151 GGACATCCCC GCCACGGTCG TCCCCACCCT GCGCCGCGAT CACGGTGACA

57201 CCACCCAGCT CACCCGTGCC GCAGCGCACG CCTTCACCGC CGGCGCCACC

57251 GTCGACTGGC GGCGCTGGTT CCCGGCCGAC CCCACCCCCC GCACGATCGA

57301 CCTGCCCACC TACGCCTTCC AGCGCCCGCAG CTACTGGTTG CCGGTGGACG

57351 GTGTCGGAGA TGTGCGGTCG GCCGGGCTGC GGCGGGTGGA ACACTCGCTG

57401 TTGCCCGCGG CGCTCGGTCT CGCCGATGGT GCGCTCGTGC TGACCGGACG

57451 GCTCGCGGCG TCCGGTGGTG GTGGCGGTTG GCTCGCGGAT CACGCGGTGG

57501 CGGGCACGAC GCTCGTCCCC GGTGCCGCGC TGGTCGAGTG GGCGTTGCGG

57551 GCCGCCGACG AGGCGGGCTG CCCCTCCCTT GAGGAGCTGA CGCTCCAGGC

57601 ACCTCTGGTG CTGCCCGGCT CCGGGGGCCT CCAGGTCCAA GTGGTCGTGG

57651 GTCCGGCCGA CGGACAGGGC GGCCGGCGTG AGGTGCGCGT CTTCTCGCGT

57701 GTCGACTCGG ACGACGAGGC AGCGGGGCAG GACGAGGGGT GGTCGTGTCA

57751 CGCGACCGGT GTGCTGAGCC CCGAGCCCGG TGCGGTACCG GACGGGCTCA

57801 GCGGACAGTG GCCCCCGACG GGCGCCGAGC CGCTGGAGAT CAGTGATCTC

57851 TACGAGCAGG CGGCATCGGC GGGATACGAG TACGGGCCGT CGTTCCGGGG

57901 CCTGCGCTCC GTGTGGCGGC ACGGGCATAA CCTGCTGGCA GAGGTGGAGC

57951 TGCCCGAACA GGCAGGTGCG CACGACGACT TCGGCATCCA CCCCGTACTG

58001 CTGGACGCCG CGCTGCACCC GGCGCTGCTG CTCGACCAGA ACGCGCCCGG

58051 CGAAGAGCAA GAGCCAGCCC AGCCCGCTCT TCGCCTGCCG TTCGTGTGGA

58101 ACGGCGTCTC CCTGTGGGCC ACCGGCGCCG CGACCGTGCG GGTACGGCTG

58151 GCCCCGCACG GGGGAGGGGA GACGGACGAT AGCGCCGGGC TGCGCGTGAC

58201 GGTCGCCGAC GCCACCGGAG CACCGGTGCT GAGCGTGGAC TCCCTCGCTC

58251 TGCGCCCCGC TGACCCCGAA CTGCTGCGCA CGGCCGGTCG GGCGGGCAGC

58301 GGCACCAACG GCTTGTTCAC GGTGGAGTGG ACCGCTCTGC CCCCGGCGGA

58351 CGTGGCCGAC CACGCCGCAG GCGACGGCTG GGCGGTGCTC GGTCAGGACG

58401 TACCCGACTG GGCCGGAGCG GACATGCCCC GGCATCCCGA CATGGCCTCC

58451 CTGTCGGCCG CGCTGGACGA GGGAACGCAG GCCCCTGCGG CCGTCTTCGT

58501 GGAGACCACA GCCACATCGC ACGCCACACC GAACACCGCA GCGGACGTGA

58551 CGCTCGACGC GTCCGGCCGG GCGGTCGCCG AGCGCACCCT GCACCTGCTG

58601 CGGGACTGGC TCGCCGAACC GCGCCTCGCC GAGACCCGGC TCGTCCTCAT

58651 CACCCACCAC GCGGTGACGA CCCCGGCGGA CGACGACGTG AACGCCGCAC

58701 CCCTCGACGT CCCGGCCGCC GCCCTGTGGG GACTGATCCG CAGCGCACAG

58751 GCCGAACACC CGGACCGCTT CGTTCTGTTG GACACCGACG CGAAGGCCAA

58801 CACCGACCCC GGCCCCGACA CCAGTACTGA CCACAGCACC GCATCGGGTA

58851 CGTACCGAAC CGTCATCGCG CGGGCCCTCG CCACCGGGGA GCCACAGCTG

58901 GCCGTGCGCG CGGGAGAACT GCTGGCTCCC CGCCTCGCCC GAGCCGCCAC

58951 CCCCACACCC GAGACCCCCA CACCCGAGAC ACAGCCCGAC ACCGGATCCG

59001 GGTCCGAGGC CGGGGCCGGG TCCGGATCTG GACCCGGCGC GACACTGGAC

59051 CCCGACGGCA CCGTCCTCAT CGCGGGCGGC ACCGGCATGA TGGGTGGTCT

59101 CGTCGCCGAA CACCTGGTCC GCGCCTGGTC GGTGCGGCAC CTCCTGCTCG

59151 TCAGCCGGCA AGGGCCCGAC GCGCCGGACG CCCGCGACCT CGCCGACCGG

59201 CTGGTCGGCC TGGGCGCGAC GGTACGGATC GTCGCGGCCG ACCTGACGGA

59251 CGGGCGGGCC ACCGCGGACC TCGTCGCGTC GGTCGACCCG GCGCACCCGC

59301 TCACCGGTGT GATCCACGCG GCCGGCGTCC TGGACGACGC CGTGGTCACC

59351 GCGCAGACCT CCGACCAGCT GGCCAGGGTG TGGGCGGCCA AGGCGTCCGT

59401 CGCCGCCAAC CTGGACGCGG CCACGTCGGA GCTGCCGCTC GGCTTGTTCC

59451 TGATGTTCTC GTCCGCCGCC GGTGTCCTCG GCAACGCGGG CCAGGCCGGT

59501 TACGCGGCCG CCAACGCCTT CGTCGACGCC CTGGTCGGCC GCCGTCGCGC

59551 CACCGGCCTG CCCGGCCTGT CGATCGCCTG GGGCCTGTGG GCGCGCGGCA

59601 GCGCCATGAC CCGGCACCTG GACGACGCCG ACCTCGCGCG GCTGCGTGCC

59651 GGCGGGGTCA AGCCCCTGCT GGACGAGCAG GGCCTCGCCC TCCTCGACGC

59701 GGCGCGCGCC ACCGCCGCGC ACACCTCGCT GGTGGTCGCG GCCGGTATCG

59751 ACGTACGCGG ACTGAACAGG GACGACGTCC CCGCGATCCT CCGCGACCTG

59801 GCGGGCCGGA CCCGCCGCAG GGCGGCCGCC GACTCCACCG TCGACCAGGC

59851 CGCGCTGGAG CGGCGCCTCA CGGGCCTGGA CGAGGCCGAG CGCCGGGCTG

59901 TCGTCACCGA CGTCGTACGC GAATGCGTGG CGGCCGTGCT CGGCCACCGG

59951 TCGGCGGCCG ACGTACGCAC CGAGGCCAAC TTCAAGGACC TCGGCTTCGA

60001 CTCGCTCACT GCGGTGCAGC TGCGCAACCG CCTCTCGGCG GCGAGCGGCC

60051 TCCGCCTGCC CGCCACCCTG GCCTTCGACC ACCCCACCCC CCAGGCGCTG

60101 GCGGCGTACC TCGGCACGCG CCTGAGCGGC CGGACCGCCA CCCCCGTCGC

60151 ACCCGTGGCG CCTTCCGCGG CCGCGACGGA CGAGCCGGTG GCGATCGTCG

60201 CGATGGCCTG CAAGTACCCG GGTGGAGCGA CCTCGCCGGA AGGCCTCTGG

60251 GACCTGGTCG CGGAGGGCGT GGACGCGGTC GGCGCCTTCC CGACGGGCCG

60301 CGGCTGGGAC CTCGAACGGC TCTTCCACCC CGACCCGGAC CACCCCGGCA

60351 CGAGTTACGC CGACGAAGGG GCCTTCCTTC CTGACGCGGG CGATTTCGAT

60401 GCGGCGTTCT TCGGGATCAA TCCGCGGGAG GCGCTGGCGA TGGATCCGCA

60451 GCAGCGGCTG TTGCTGGAGG CGTCGTGGGA GGTGTTGGAG CGTGCGGGTA

60501 TCGACCCGAC GACGCTCAAG GGCACCCCGA CCGGCACGTA CGTCGGCGTG

60551 ATGTACCACG ACTACGCGGC AGGCCTCGCC CAGGACGCCC AACTGGAGGG

60601 CTACTCCATG CTCGCCGGCT CCGGCAGCGT GGTGTCCGGC CGCGTCGCCT

60651 ACACCCTGGG GCTTGAGGGT CCTGCGGTGA CGGTCGACAC CGCGTGCTCC

60701 TCGTCCCTGG TCTCCATCCA CCTGGCCGCG CAAGCACTGC GACAGGGCGA

60751 GTGCACTCTC GCCCTCGCGG GCGGCGTGAC CGTCATGGCC ACGCCCGAGG

60801 TGTTCACCGG ATTCTCGCGC CAGCGCGGCC TGGCCCCCGA CGGCCGCTGC

60851 AAGCCGTTCG CCGCCGCCGC CGACGGCACC GGCTGGGGCG AGGGTGTCGG

60901 TGTGTTGTTG CTCGAGCGGT TGTCGGATGC GCGGCGTCAT GGGCGTCGGG

60951 TGTTGGGTGT GGTGCGGGGT TCGGCGGTGA ATCAGGACGG TGCGAGTAAT

61001 GGGTTGACGG CGCCGAATGG TCCGTCGCAG GAGCGGGTGA TTCGTCAGGC

61051 GTTGGCCAGT GGTGGGTTGT CGTCGGTGGA TGTTGATGTG GTGGAGGGGC

61101 ATGGGACGGG GACCACGTTG GGTGATCCGA TCGAGGCGCA GGCTCTGCTG

61151 GCCACGTATG GGCAGGGGCG TCCGGTGGAT CGTCCGTTGT GGTTGGGGTC

61201 GGTGAAGTCG AATATTGGTC ATACGCAGGC GGCTGCGGGT GTTGCGGGTG

61251 TCATCAAGAT GGTGATGGCG ATGCGGCATG GTGTGGTGCC GGCGAGTTTG

61301 CATGTGGATG TGCCGTCGCC GCATGTGGAG TGGGATTCGG GTGCGGTGCG

61351 GTTGGCGGTT GAGTCGGTGC CATGGCCGGA GGTGGAGGGT CGTCCGCGTC

61401 GGGCGGGTGT GTCGTCGTTC GGGGCTTCGG GAACGAATGC GCACGTGATC

61451 GTGGAGTCTG TGCCCGATGG GCTGGGGGAG GACTCGGTAT CGGTCAGTGG

61501 TGAGGCTCCC GAGACTGAGA CTGACGGGCG CTTGGTGCCG TGGGTGGTAT

61551 CGGCCCGCAG CCCGCAGGCC CTGCGCGACC AGGCACTACG CCTGCGTGAT

61601 GCGGTGGCGG CCGACTCAAC GGTGTCGGTG CAGGATGTGG GCTGGTCGCT

61651 GCTGAAGACG CGTGCGCTGT TCGAGCAGCG GGCGGTGGTG GTGGGGCGTG

61701 AGAGGGCTGA ACTCCTGTCG GGGCTTGCTG TGTTGGCCGC TGGCGAGGAG

61751 CACCCGGCTG TGACGCGGTC CCGTGAGGAC GGGGTTGCTG CGAGCGGTGC

61801 TGTGGTGTGG CTGTTCAGTG GTCAGGGCAG TCAGTTGGTC GGTATGGGTG

61851 CTGGTTTGTA TGAGCGGTTC CCGGTGTTTG CGGCTGCGTT TGATGAGGTG

61901 TGCGGCCTGT TGGAGGGGCC GTTGGGCGTG GAGGCGGGTG GGTTGCGGGA

61951 GGTGGTGTTC CGTGGCCCGA GGGAGCGGTT GGATCACACG ATGTGGGCGC

62001 AGGCGGGGTT GTTTGCGCTG CAGGTGGGGT TGGCCCGGTT GTGGGAGTCG

62051 GTCGGGGTGC GGCCGGATGT GGTGCTCGGG CATTCGATCG GTGAGATCGC

62101 GGCCGCGCAT GTGGCGGGGG TCTTTGATCT GGCGGATGCC TGTCGGGTGG

62151 TGGGGGCGCG GGCCCGTTTG ATGGGTGGGC TGCCTGAGGG CGGGGCGATG

62201 TGCGCGGTGC AGGCCACGCC CGCCGAGCTG GCCGCCGACG TGGACGACTC

62251 TGGTGTGAGT GTGGCGGCGG TCAACACACC TGATTCGACG GTGATTTCAG

62301 GGCCGTCTGG TGAGGTGGAT CGGATTGCTG GGGTGTGGCG GGAGCGTGGG

62351 CGTAAGACGA AGGCGCTGAG CGTCAGTCAT GCCTTCCACT CGGCGTTGAT

62401 GGAGCCGATG CTCGCGGAGT TCACCGAAGC GATACGAGAG GTCAAGTTCA

62451 CGCGGCCGAA GGTGTCGTTG ATCAGCAACG TCTCTGGTCT GGAGGCGGGT

62501 GAGGAGATCG CGTCCCCGGA GTACTGGGCA CGCCATGTAC GCCAGACAGT

62551 GCTCTTCCAG CCCGGCATCG CCCAAGTGGC TTCCACGGCA GGCGTGTTTG

62601 TCGAGCTCGG CCCCGGCCCC GTACTGACTA CTGCCGCCCA GCACACCCTG

62651 GACGACGTAA CCGATAGGCA TGGCCCCGAA CCGGTACTGG TGTCCTCGCT

62701 GGCCGGTGAG CGTCCTGAGG AGTCGGCGTT CGTGGAGGCG ATGGCTCGTC

62751 TGCATACCGC TGGTGTTGCT GTGGACTGGT CGGTGTTGTT CGCGGGTGAT

62801 CGTGTGCCTG GGCTGGTGGA GTTGCCGACG TATGCGTTCC AGCGGGAGCG

62851 GTTCTGGTTG AGCGGCCGTT CTGGGGGTGG GGATGCGGCG ACTTTGGGTC

62901 TGGTGGCGGC GGGGCATCCG TTGTTGGGTG CGGCGGTGGA GTTCGCGGAC

62951 CGGGGTGGGT GTCTGCTGAC CGGTCGGCTG TCGCGGTCTG GGGTGTCGTG

63001 GCTTGCTGAT CATGTGGTGG CGGGTGCGGT TTTGGTGCCG GGTGCTGCGT

63051 TGGTGGAGTG GGCGTTGCGG GCCGGTGATG AGGTCGGTTG TGTGACGGTG

63101 GAGGAGTTGA TGTTGCAGGC GCCTTTGGTG GTGCCTGAGG CGTCGGGTCT

63151 GCGGGTTCAG GTGGTGGTCG AGGAGGCGGG TGAGGACGGG CGGCGCGGTG

63201 TCCAGATCTA TAGCCGGCCT GACGCGGACG CCGTGAGCGG CGACGACTCG

63251 TGGATCTGCC ACGCGACCGG CACCCTCACC CCCCAGCACA CCGACGCTCC

63301 GAACGACGGA CTGGCCGGCG CGTGGCCCGC GGCGGGCGCC GTGCCGGTGG

63351 ACCTGGCGGG CTTCTACGAG CGCGTGGCGG ACGCGGGCTA TGCGTACGGC

63401 CCGGGGTTCC AGGGGCTGCG TGCCGTGTGG CGGCACGGTC AGGACCTGCT

63451 GGCCGAGGTC GTCCTGCCCG AAGCCGCGGG TGCCCATGAC GGCTACGGCA

63501 TCCACCCCGC CCTCCTCGAC GCCACCCTCC ACCCGGCCCT GCTCCTCGAC

63551 TGGCCCGGGG AGGTGCAGGA CGACGACGGG AAGGTCTGGC TGCCTTTCAC

63601 CTGGAACCAG GTCTCCTTGC GGGCTGCGGG AGCCGCCACC GTACGCGTAC

63651 GTCTCTCGCC CGGCGAGCAC GACGAGGCGG AACGGGAAGT ACAGGTACTG

63701 GTGGCCGACG CCACCGGGAC CGACGTCCTG AGCGTGGGGT CGGTGACGTT

63751 GCGTCCCGCC GACATCCGGC AACTGCAGGC CGTGCCGGGT CACGACGACG

63801 GTCTGTTCTC GGTGGACTGG ACGCCGCTGC CGCTGTCGCG GACGGATGTG

63851 TCGCAGACGG ATGCCGACGG GGATGCCGAC TGGGTTGTGC TCTCGGACGG

63901 TGTCGGCAGC CTGGCTGATG TGGTGTCGGC GGCGGGTGGT GAAGCGCCGT

63951 GGGCAGTGGT CGCTCCCGTC GGTGCATCCG CGGGCGGCGG CCTTGCCGGC

64001 TTTGACCGCC GTGAGGGTCT TGACGGTCGG CTGGTCGTGG AGCGGGTGTT

64051 GTCACTCGTA CAGGAGTTCC TGGCCGCGCC GGAGCTGGCC GAGTCCCGGC

64101 TCCTCGTGCT GACCCGCGGC GCCGTGGCGA CCGGCGGCGA CGGCGACGGT

64151 GATGTGGACG CGTCCGCCGC AGCCGTATGG GGCCTGGTCC GCAGTGCTCA

64201 GTCCGAGAAC CCGGGCCGCT TCATCCTGCT CGACGTGGAC ATGGACGTGG

64251 ACGTCGACGT GGACATGGAC GTGGACGTCG ACGTGGACGT CGACGTGGAC

64301 GTGGACGGAG ACGGCAATGG CAGCGACCTG GACCCGGACC TGAACGGCCG

64351 ACGACTTCCC CACGCCACCC TGCGTCACGC CGCCGAGGAA CTCGACGAGC

64401 CCCAACTCGC CCTGCGCGAC GGACAACTGC TCGTTCCGCG GCTGGTCCGC

64451 GCCACCGGCG GCGGACTCGT CGTGGCGCCC ACCGACCGTG CCTGGCGCCT

64501 GGACAAGGGA AGCGCCGAGA CGCTGGAGAG CGTCGCGCCG GTCGCGTACC

64551 CCGGAGTCAT GGAACCCCTG GGCCCCGGCC AGGTCCGCCT CGGCATCCAC

64601 GCCGCGGGCA TCAACTTCCG CGACGTCCTG GTCAGCCTCG GCATGGTGCC

64651 CGGCCAGGTC GGCCTGGGCG GCGAAGGCGC CGGTGTCGTG ACGGAGACAG

64701 GCCCCGATGT CACCCACCTG TCGGTCGGCG ACCGCGTGAT GGGCGTCCTC

64751 CACGGCTCCT TCGGCCCGAC GGCCGTGGCG GACACCCGCA TGGTCGCGCC

64801 GGTTCCGCAG GGCTGGGACA TGCGGCAGGC GGCCGCGATG CCCGTCGCGT

64851 ATCTGACGGC TTGGTACGGG TTGGTGGAGC TGGCCGGTCT GAAGGCGGGC

64901 GAGCGCGTGC TGATCCACGC AGCCACGGGT GGTGTGGGAA TGGCGGCGGT

64951 GCAGATCGCC CGTCACCTGG GTGCCGAGGT GTTCGCCACC GCCAGTGCAG

65001 CCAAGCACGT CGTACTGGAA GAGATGGGCA TCGACGCCGC CCACCGCGCC

65051 TCCTCCCGGG ACCTCGCCTT CGAGGACACC TTCCGGCAGG CCACCGACGG

65101 GCGCGGCATG GACGTCGTCC TCAACAGCCT GACCGGCGAG TTCATCGACG

65151 CATCTCTGCG GTTGCTCGGC GACGGCGGCC GGTTCCTGGA GATGGGCAAG

65201 ACCGATGTGC GCACGCCGGA GGAGGTGGCC GCGGAGTACC CGGGTGTCAC

65251 CTACACCGTG TACGACCTCG TCACCGACGC GGGGCCGGAT CGCATCGCGG

65301 TCATGATGAG TGAGCTGGGC GAGAGGTTCG CTTCCGGTGC CCTTGACCCT

65351 CTGCCGGTGC GTTCCTGGCC GCTGGACAAG GCGCGTGAGG CGTTCCGGTT

65401 CATGAGTCAG GCCAAGCACA CCGGCAAACT CGTACTCGAC GTGCCCGCAC

65451 CGCTCGACCC CGACGGGACC GTCCTGATCA CCGGAGGCAC GGGGGCGCTG

65501 GGGCAGGTCG TGGCCGAGCA TCTGGTGCGG GAGTGGGGCG TACGGCACCT

65551 GCTGCTGGCC AGCCGCCGTG GACTGGACGC CCCCGGCAGC GGTGAACTCG

65601 CCGACAGGCT GTCGGACTTG GGCGCCGAGG TGACCGTCGC GGCGGCCGAT

65651 GTGAGCGACC CGGCCTCGGT GGTGGAGCTG GTCGGCAAGA CGGATCCCTC

65701 GCATCCGTTG ACGGGTGTCG TGCACGCGGC GGGCGTGCTT GAGGACGGGA

65751 TCGTGACGGC TCAGACGCCT GAGGGGCTGG CGCGGGTGTG GGCGGCCAAG

65801 GCCGCTGCGG CGGCGAATCT CCATGAGGCG ACCCGGGAGA TGCGTCTCGG

65851 TCTGTTCGTG GTGTTCTCCT CGGCGGCCGC CACGCTCGGC AGTCCGGGCC

65901 AGGCCAACTA CGCGGCTGCC AATGCCTATT GTGACGCGCT GATGCAGCGC

65951 CGACGGGCGG CGGGCCAGGT CGGCCTGTCG GTCGGCTGGG GTCTCTGGGA

66001 GGCACCGGAC GCCAAGCCGG GTGTTGCCGC CGACGCCAAA CCGGATGTTG

66051 CCGCCGACGC CAAGACGGGA GTTGCCGCCG ACGGCACTCC CCAGGGCATG

66101 ACCGGCACCC TGAGCGGCAC CGACGTGGCC CGCATGGCAC GCATCGGCGT

66151 CAAGGCGATG ACCAGCGCAC ACGGTCTCGC CCTGCTCGAC GCCGCACACC

66201 GCCACGGCCG CCCCCACCTC GTCGCCGTCG ACCTCGACAC CCGCGTCCTG

66251 GCGCACAAAC CCGCCCCGGC CCTCCCCGCC CTCCTGCGCG CCTTCGCCGG

66301 AGACCAGGGA GGCCAGGGAG GCGGCCGAGG CGGCGGTCGG GGCGGCGGCC

66351 CGGCACGACC GGCGGCCGCC ACCACCCGGC AGAACGTCGA CTGGGCCGCG

66401 AAGCTCTCCG TCCTGACAGC CGAGGAACAG CACCGCACCC TCCTCGACCT

66451 GGTACGGACG CACGCGGCAG CCGTCCTCGG GCACGCGGGC ACCGACGCCG

66501 TACGCGCCGA CGCCGCCTTC CAGGATCTCG GCTTCGACTC CCTCACCGCG

66551 GTCGAACTGC GCAACCGCCT CTCCGCCTCC ACCGGCCTGC GCCTGCCCGC

66601 CACGTTCATC TTCCGGCACC CGACCCCGTC GGCCATCGCC GACGAACTGC

66651 GCGCACAGCT GGCCCCCGCG GGGGCCGACC CGGCCGCGCC GCTCTTCGGT

66701 GAACTGGACA AGCTGGAGAC GGTGATCACG GGGCACGCGC ACGACGAGAG

66751 CACCCGGACC CGCCTGGCGG CACGCCTGCA GAACCTGCTG TGGCGCCTGG

66801 ACGACACTTC GGCCCGCTCG GACCACGCGG CCGGCGCGAG CGACGCCGAC

66851 GGCGACGCCG TCGAGAACCG AGACCTCGAG TCCGCGTCGG ACGACGAGCT

66901 CTTCGAGCTG ATCGACCGAG AACTGCCTTC TTGATCAGGA GTGGAGAAGA

66951 CATGCCGGGT ACGAACGACA TGCCGGGTAC CGAGGACAAG CTCCGCCACT

67001 ACCTGAAGCG AGTGACCGCG GATCTCGGAC AGACCCGTCA GCGCCTGCGC

67051 GACGTGGAGG AGCGCCAGCG GGAACCGATC GCCATCGTCG CGATGGCCTG

67101 CCGCTACCCG GGCGGGGTGG CCTCCCCCGA GCAGCTGTGG GACCTGGTCG

67151 CCTCACGCGG CGACGCCATC GAGGAGTTCC CCGCCGACCG CGGCTGGGAC

67201 GTGGCGGGCC TCTACCACCC CGACCCGGAC CACCCCGGCA CGACCTATGT

67251 ACGAGAGGCC GGATTCCTGC GGGACGCCGC CCGCTTCGAC GCCGACTTCT

67301 TCGGCATCAA CCCGCGCGAG GCGCTCGCCG CCGACCCGCA GCAACGGGTG

67351 CTCCTCGAAG TGTCGTGGGA ACTGTTCGAG CGGGCGGGCA TCGACCCCGC

67401 CACGCTCAAG GACACCCTCA CCGGCGTGTA CGCGGGGGTG TCCAGCCAGG

67451 ACCACATGTC CGGGAGCCGG GTCCCGCCGG AGGTCGAGGG CTACGCCACC

67501 ACGGGAACCC TCTCCAGCGT CATCTCCGGC CGCATCGCCT ACACCTTCGG

67551 CCTGGAGGGC CCGGCGGTGA CGCTCGACAC GGCGTGCTCG GCATCGCTGG

67601 TCGCGATCCA CCTCGCCTGC CAGGCCCTGC GCCAGGGCGA CTGCGGCCTG

67651 GCGGTGGCGG GAGGCGTGAC CGTACTGTCC ACGCCGACGG CGTTCGTGGA

67701 GTTCTCACGC CAGCGCGGAC TCGCACCGGA CGGCCGCTGC AAGCCGTTCG

67751 CCGAGGCCGC CGACGGCACC GGATTCTCCG AGGGCGTCGG CCTGATCCTC

67801 CTGGAACGCC TCTCCGACGC CCGCCGCAAC GGACATCAAG TACTCGGCGT

67851 CGTACGCGGA TCGGCCGTCA ACCAGGACGG CGCGAGCAAC GGCCTGACCG

67901 CCCCGAACGA CGTCGCCCAG GAACGCGTGA TCCGCCAGGC CCTGACCAAC

67951 GCCCGCGTCA CCCCGGACGC CGTCGACGCC GTGGAGGCAC ACGGCACCGG

68001 CACCACGCTC GGCGACCCGA TCGAGGGGAA CGCACTCCTC GCGACGTACG

68051 GAAAGGACCG CCCCGCCGAC CGGCCGCTGT GGCTCGGCTC TGTGAAGTCG

68101 AACATCGGCC ACACGCAGGC GGCTGCGGGC GTCGCAGGCG TCATCAAGAT

68151 GGTGATGGCG ATGCGCCACG GCGAGCTGCC CGCCTCCCTG CACATCGACC

68201 GGCCCACGCC CCACGTGGAC TGGGAGGGCG GGGGAGTGCG GTTGCTCACC

68251 GATCCCGTGC CGTGGCCACG GGCCGACCGC CCCCGCCGCG CGGGGGTCTC

68301 CTCCTTCGGC ATCAGCGGCA CCAACGCCCA CCTGATCGTG GAACAGGCCC

68351 CCGCCCCGCC CGACACGGCC GACGACGCCC CGGAAGGCGC CGCAACCCCC

68401 GGCGCTTCCG ACGGCCTCGT GGTGCCGTGG GTGGTGTCGG CCCGTAGTCC

68451 GCAGGCCCTG CGTGATCAGG CCCTGCGTCT GCGCGACTTT GCCGGTGACG

68501 CGTCCCGAGC GCCGCTCACC GACGTGGGCT GGTCTTTGCT GCGGTCGCGT

68551 GCGCTGTTCG AGCAGCGGGC GGTGGTGGCG GGGCGTGAGA GGGCTGAACT

68601 GCTGGCGGGG CTGGCTGCGT TGGCCGCTGG TGAGGAGCAC CCGGCTGTGA

68651 CGCGGTCCCG TGAGGAAGCG GCGGTTGCTG CGAGCGGTGA TGTGGTGTGG

68701 CTGTTCAGTG GTCAGGGCAG TCAGTTGGTC GGTATGGGTG CTGGTTTGTA

68751 TGAGCGGTTC CCGGTGTTTG CGGCTGCGTT TGATGACGTG TGCGGCTTGC

68801 TGGAGGGGGA GCTGGGGGTT GGTTCGGGTG GGTTGCGGGA GGTGGTGTTC

68851 TGGGGCCCGC GGGAGCGGTT GGATCACACG GTGTGGGCGC AGGCGGGGTT

68901 GTTTGCGTTG CAGGTGGGGT TGGCCCGGTT GTGGGAGTCG GTCGGGGTGC

68951 GGCCGGATGT GGTGCTCGGG CATTCGATCG GTGAGATCGC GGCCGCGCAT

69001 GTGGCGGGGG TCTTTGATCT GGCGGATGCG TGTCGGGTGG TGGGGGCGCG

69051 GGCGCGTTTG ATGGGTGGGT TGCCTGAGGG TGGGGCGATG TGTGCGGTGC

69101 AGGCCACGCC CGCCGAGCTG GCCGCGGATG TGGATGGCTC GTCCGTGAGT

69151 GTGGCGGCGG TCAACACACC TGACTCGACG GTGATTTCAG GTCCGTCGGG

69201 TGAGGTGGAT CGGATTGCTG GGGTGTGGCG GGAGCGTGGG CGTAAGACGA

69251 AGGCGCTGAG CGTGAGTCAT GCTTTCCATT CGGCGTTGAT GGAGCCGATG

69301 CTCGGGGAGT TCACGGAAGC GATACGAGGG GTCAAGTTCA GGCAGCCGTC

69351 GATCCCGCTC ATGAGCAATG TCTCCGGAGA GCGGGCCGGC GAGGAGATCA

69401 CATCCCCGGA GTACTGGGCG AGGCATGTAC GCCAGACAGT GCTCTTCCAG

69451 CCCGGCGTCG CCCAAGTGGC CGCTGAGGCA CGCGCGTTCG TCGAACTCGG

69501 CCCCGGCCCC GTACTGACCG CCGCCGCCCA GCACACCCTC GACCACATCA

69551 CCGAGCCGGA AGGCCCCGAG CCGGTCGTCA CCGCGTCCCT CCACCCCGAC

69601 CGGCCGGACG ACGTGGCCTT CGCGCACGCC ATGGCCGACC TCCACGTCGC

69651 CGGTATCAGC GTGGACTGGT CGGCGTACTT CCCTGACGAC CCCGCCCCCC

69701 GCACCGTCGA CCTGCCCACC TACGCCTTCC AGGGGCGGCG CTTCTGGCTG

69751 GCGGACATCG CGGCGCCCGA GGCCGTGTCC TCGACGGACG GTGAGGAGGC

69801 CGGGTTCTGG GCCGCCGTCG AAGGTGCGGA CTTCCAGGCG CTCTGCGACA

69851 CCCTGCACCT CAAGGACGAC GAGCACCGCG CGGCTCTGGA GACGGTGTTC

69901 CCCGCGCTGT CCGCGTGGCG GCGCGAACGA CGTGAGCGGT CGATCGTCGA

69951 TGCCTGGCGG TACCGGGTCG ACTGGCGGCG CGTCGAGCTG CCGACACCCG

70001 TTCCGGGCGC CGGTACCGGT CCCGACGCCG ACACGGGCCT CGGGGCGTGG

70051 CTGATCGTGG CTCCCACGCA CGGGTCGGGT ACTTGGCCGC AAGCCTGTGC

70101 CCGGGCGTTG GAGGAGGCGG GCGCGCCGGT ACGTATCGTC GAGGCCGGCC

70151 CGCACGCCGA CCGGGCGGAC ATGGCGGACC TGGTCCAGGC ATGGCGGGCA

70201 AGCTGTGCGG ACGACACCAC CCAGCTCGGA GGAGTGCTCT CCCTGCTGGC

70251 TCTCGCCGAG GCACCGGCCA CCAGTTCCGA CACCACTTCC CACACCAGTA

70301 CCAGTTGCGG TACCGGCTCT CTCGCGTCCC ACGGCCTCAC CGGCACCTTG

70351 ACGCTGCTGC ACGGTCTGCT GGATGCGGGC GTCGAAGCGC CTCTCTGGTG

70401 TGCCACGCGC GGCGCCGTGT CGTGCGGCGA CGCCGATCCG CTCGTCTCCC

70451 CGTCGCAGGC CCCGGTCTGG GGACTCGGAC GCGTGGCCGC CCTGGAGCAT

70501 CCGGAGTTGT GGGGCGGCCT GGTCGACCTG CCCGCCGACC CGGAGTCGCT

70551 CGACGCGAGC GCGTTGTATG CGGTTCTGCG CGGAGACGGC GGCGAGGATC

70601 AGGTCGCGCT GCGCCGGGGC GCGGTCCTCG GCCGTCGCCT GGTGCCCGAC

70651 GCAACCCCGG ACGTGGCCCC CGGCTCGTCC CCGGACGTGT CCGGAGGCGC

70701 AGCCCATGCC GACGCGACCT CCGGGGAGTG GCAGCCGCAT GGTGCCGTCC

70751 TCGTCACCGG AGGCGTCGGC CACCTGGCCG ATCAGGTCGT ACGGTGGCTC

70801 GCCGCGTCCG GCGCCGAACA CGTCGTACTC CTGGACACGG GCCCCGCCAA

70851 CAGCCGTGGT CCCGGCCGGA ACGACGACCT CGCCGCGGAA GCCGCCGAAC

70901 ACGGCACCGA GCTGACGGTC CTGCGGTCCC TGAGCGAGCT GACAGACGTA

70951 TCCGTACGTC CCATACGGAC CGTCATCCAC ACATCGCTGC CCGGCGAGCT

71001 CGCGCCGCTG GCCGAGGTCA CCCCCGACGC GCTCGGCGCG GCCGTGTCCG

71051 CCGCCGCGCG GCTGAGCGAA CTCCCCGGCA TCGGGTCAGT GGAGACCGTG

71101 CTGTTCTTCT CCTCCGTGAC GGCTTCGCTC GGCAGTAGGG AGCACGGCGC

71151 GTACGCCGCC GCCAACGCCT ACCTCGACGC CCTGGCGCAA CGGGCCGGTG

71201 CCGATGCTGC GAGCCCCCGG ACGGTCTCGG TCGGGTGGGG CATCTGGGAT

71251 CTGCCGGACG ACGGTGACGT GGCACGCGGC GCCGCCGGGC TGTCCCGGAG

71301 GCAGGGACTC CCGCCGCTGG AACCGCAGTT GGCGCTCGGC GCCCTGCGCG

71351 CGGCGCTCGA CGGGGGCAAG GGGCACACGC TGGTCGCCGA CATCGAGTGG

71401 GAGCGGTTCG CGCCGCTGTT CACGCTGGCC AGGCCCACCC GGCTGCTCGA

```
71451  CGGGATCCCC GCGGCCCAGC GGGTCCTCGA CGCCTCCTCG GAGAGCGCCG
71501  AGGCCTCGGA GAACGCCTCG GCCCTCCGTC GCGAACTGAC GGCCCTGCCC
71551  GTGCGGGAGC GGACCGGGGC ACTTCTCGAC CTGGTCCGCA AACAGGTGGC
71601  CGCCGTCCTG CGCTACGAGC CGGGCCAAGA CGTGGCGCCC GAGAAGGCCT
71651  TCAAGGACCT GGGCTTCGAC TCGCTCGTGG TCGTGGAGCT GCGCAACCGG
71701  CTGCGCGCCG CCACCGGGCT CCGGCTGCCC GCCACCCTGG TCTACGACTA
71751  CCCCACACCC CGCACCCTCG CCGCACACCT GCTGGACAGG GTGCTGCCCG
71801  ACGGCGGCGC GGCAGAGCTC CCCGTGGCCG CCCACCTGGA CGACCTGGAG
71851  GCGGCCCTCA CCGACCTGCC GGCCGACGAC CCCCGGCGCA AGGGCCTGGT
71901  CCGGCGTCTA CAGACGCTGC TGTGGAAGCA GCCCGACGCC ATGGGGGCGG
71951  CGGGCCCCGC CGACGAGGAG GAGCAAGCCG CGCCCGAGGA CCTGTCGACC
72001  GCGAGCGCCG ACGACATGTT CGCCCTGATC GACCGGGAGT GGGGCACGCG
72051  GTGAGCGGGG TGGAGCGGGG TGTGGGGTCG GCGGGCCCTG TGGAACAGGG
72101  TGACGGACTC GCGGGCCTGG TCGAGCGGGC CGAGGCGCTG GCCGCTCTGC
72151  GGGGCGCCTT CGACGGCTCC CCGGGCACCG GCGGCAGCCT CGTCGTGCTC
72201  AGCGGCGCGG TGGGCACCGG CAAGACCGCG CTGCTACGGG CGTGGGCCGA
72251  CCGCATCGGC GCCGATGCCG ACGCCCTGGT CCTGACCGCC ACCGCCTGCC
72301  GCGCCGAGCG CGACCTGCCG CTTGGCGTCC TGGAACAGCT GGTACGCAGC
72351  CCCGGCCTGC CCCCGGCCAG CGCCGAGCGC GCGCTGGCGT GGTGGGACGA
72401  GGAGGCCTCG GCCACCCCCG GAAAGACGGA CGCGAACGGG ACGAGTGCCA
72451  ACGGGACGGA CGCCAACGGG ACGGGCGCGG GACAGACGGG CGCGGGGCAG
72501  GCGGGCGTGG GACAGACGGG CGTGGGCGGA GAGCCCGTCC TGGCCGCCTC
72551  CGCCCTGCGA GGCCTGTGCG AGGTGCTGCG GGACCTGCTC GCCGAGCGGC
72601  CCGTCGTGGT CGCCGTCGAC GACGCGCACC ATGCCGACGC GGCGTCGCTC
72651  CAGTGCCTGC TCTCCGTGGT GCGCCGGCTG CGGTCGGCAC GACTCCATGT
72701  GCTGTTCACC GAGTACGCCC ATCAGAAGGC GCAGAACGCC CTGCTGAGCA
```

72751 GCGAGTTCCT GCACGAGCCC GCCCTGCGGC GGATCCGCCT GGAACCGCTG

72801 TCGAAGGCGG GCGTGGAGGC CTTGCTCGCC CGGCACCTCG ACGAGCGGAC

72851 GGCACAAGAC CTCACCCCCG TCGTCCACGG CATGAGCGCG GGCCACCCGC

72901 TCCTCGTACG GGCGCTGGCC GAGGACCACC GTGCGGCGGG CGGCGCCGGG

72951 GAGGCGTACG GTCGTGCCGT CCTCAGCTTT CTGTACCGGC ACGAGACTCC

73001 GGTCACCCAA GTCGCCCGCG CCATCGCTGC GTTGGGCGCG CACGCCGGAC

73051 CCGGTCAGGT CGGGCGGCTG CTCGATGTCG ACGCGGCGTC CGTCGAGCGG

73101 GCCGTGCGGC AGCTGACCGT CGCGGAGGTG CTGCACGAGG GCCGCCTGTG

73151 CCACCCGGCG TTCGCGGCGG CGGTCCTGGA CGGCATGCCG CCCGAGGAAC

73201 GCCGCGCCCT GCACGGACGG GTCGCCGACC TCCTGCACGA GGAGGGGGCG

73251 CCGGCCACCG AAGTGGCCGC CCACCTCGTC GCCGCCGACC GGTCCGACGC

73301 CCCGTGGGCG GTACCCGTCT TCCAGGAAGC GGCCCAACTC GCCCTGGACG

73351 AGGACCAGGT GGAGACCGGC GTCGACTATC TGCGCGCGGC CCACCAGCGG

73401 TGCCGGGGCG CCGCGCAGCG TGCCGCGGTC GTCGGTGCGC TCGCCGACGC

73451 CGAGTGGCGG CTCGACCCAG CAAAGGTCCT GCGCCACCTG CCCGACCCTG

73501 CAGCCATGGC CCCACAAACG GACCCTGCCG CCCTGGCCCC ACACACGGAC

73551 CCCGCACCCA CAGCCGCACC CACAGCCGCC CCCACCCCCA CCCCCATCCC

73601 GACCACCCCA CCCCTCCCCA CCCACCTGCT CTGGCACGGG CGGGTCGAGG

73651 AAGGCCTGGA CGCCATCGGC ACGCTCACCG GGCCCGGACC CAACCCGGCG

73701 GGTGCGCCGC CGATGAACCC CGCGGACCTG GACACCCCAT GGCTGTGGGG

73751 CGCCTACCTC TATCCCGGGC ACGTCAAGGA GCGCCTGGGA TCCGGCGCCC

73801 TGTCCCCGCA GCGCTCGACC CCGCCGGCGG TCACGCCGGA GCTCCAAGGC

73851 GCGGGCACGC TGATGAACGA CCTGCTGCAC GGCGGCGAAC GCGACGCCAC

73901 CGAGGCCGCC GAGCGCGCCC TCAACCGCTA CCGGCTCGGC CCCCGCACCA

73951 TCGCGGTCCA GACGGCCGCG CTGGCCGCCC TCACCTACCG CGACCGGCCG

74001 CACCGCGCGG CCGCCTGGTG CGACGGCCTC GTCGCCCAGG CCGACGAGCG

74051 CAACAGCCCC ACCTGGCGGG CCCTGTTCAC CGCGTGGCGT GCCCTGCTCC

74101 ACCTGCGGCA GGGCGACCCG GCCGCAGCGG AACAGCGCGC CGAAACCGCC

74151 CTCGCCCTGC TCGGATCGAA GGGCTGGGGC GCCGCGATCG GCCTGCCGCT

74201 GGCAGCCGCC GTACAGGCCA AGGCGGCCCT CGGCGATGTC GACGGGGCGG

74251 CGGCCCTCCT GGAACGGCCC GTGCCCCAGG CGGTCTTCCA GACCCGCACC

74301 GGACTGCACT ACCTGGCGGC CCGGGGCCGC TATCACCTCG CCACCGGCTG

74351 CCACTACGCC GCACTGTGCG ACTTCTACGC CTGCGGGACC CGCATGAGCA

74401 GCTGGGGAGT GGACCTGCCC GCGCTGGAGC CGTGGCGCCT CGGCGCGGCG

74451 GAAGCGTACC TGGCCCTCGG CGAAGGACTC CTGGCACGCC AACTCGTCGA

74501 CGGCCAGCTG CCGTTGCCCA CGCCTGACGA CGGCCGCACC TGGGGCATGA

74551 CGTTGCGCCT GCGGGCGGCC ACGTCCCCCG CGCCGGCCCG GGCCGAACTC

74601 CTCGACGAGG CCGTGGCGGT GCTCCGGGAG AGCGGCGACA CCTTCGAGCT

74651 GGCGCGGGCC GTCGCCGACC AGGCTGTTGC CGTACGCGAA GGGGGCGAGG

74701 CGGAACGCGC CCGGCTGCTG GCCCGCAAGG CGGAGCTGCT GGCCCGGCGC

74751 TGGGGCAGCG CCCCCGCGCC CGCCACCGTC CCCGAACCGC CGGAGCGGCC

74801 AGGACCGGCC ACTCCGGACG CCGAACTGAC CAGTGCGGAG CGGAGGGTGG

74851 CCGAGCTGGC CGCCGAAGGG TTCACCAACC GGGAGATCTC CCGGAAGCTG

74901 TGCGTCACGG TCAGCACCGT GGAACAGCAC CTGACCCGGA TCTACCGGAA

74951 GCTCGACGTC AGGCGACTGG ACCTCCAGGC AGCCCTCGGC TGACCTTCAG

75001 GCGGCCCTCG GCTGACCGCA GGCCACGCGC CTACGGTCAG CCTTCCTGAG

75051 TCAGGACCGT ACAGCCGCCG TAGGTGTAGG TGTAGGCGTG GGCGAGATCG

75101 TCGCCGCGTC CAGACCCACC ACGGCCAGCT CCTCCGGAAG GAACGGGGGA

75151 GCGGTCAGCT CCGGGAGGCG TTCGTCGGCG CGCATCGCCA TCAGGAAACG

75201 GTTGGAGCCC AGTTCGGCCT GCGGCGCGTT GAGGCTCATC ACGTCCGTGA

75251 CGATCTCGGA CGCCTTCGGG GAACGGATCG ACGCCGCGGT GATGGCCTCG

75301 GCGAACCGCA GACGCTGCTC GGTGTCCACA CCGATGAGCC GCGGATCCGT

75351 CGCCGAGACA CGGCAGTTGA CGTAGTCGAT GTCCTTGGTC GCGGCGAGGA

75401 TCCACGGGTC GTCCACGGCC GCGCCGATCG CCTTCTGCAG GGCGCGGGTG

75451 CCGGCGCGGG CGGACCCCGT ACCCTCCTGC ACGCTCCGCT CGAACTCGCG

75501 GTCGATCGTG GTGGCGCAGC GCGCGGCCGA GCTCATGCCG TGGCCGTAGA

75551 TCGGGTTGAA AGCGGTCAGC GAGTCGCCGA TGACGAGCAG ACCGTCGGGC

75601 CACTGTTCGA GGCGCTCCGG ATAGAGGCGG CGGTTGGCGC CGGAGCGGGA

75651 ACCGAAGACG GGGGTGAGTG GTTCGGCGTC CCGGAGCAGG TCGGCGAGGA

75701 TCGGGTGGTT CAGGTTCTCG GCGAAGGGGA TGAACTCGTC CTCGTGTGTG

75751 GGCAGTTGCG CGCCCCGCGT GCAGGAGAGC GTCGCGAGCC AGCGGCCGCC

75801 CTCGATGGGG TAGACCACGC CGAAGCGGCC GGGTTCGCGC ACCCGGTCGT

75851 CGGCGGCGAT GTTCACGGCG GGGAAGTGCG TCGTAGCGCC CGGCGGGGCC

75901 TTGAAGAGCC GGGTGGCGTA GGCGACGCCC GCGTCCACGA CGTCTTCCTC

75951 CAGTGCCGGC ACGCCGAGGG CGGCGAGCCA CTGCTTGAGG CGGGAGCCGC

76001 GCCCGGTGGC GTCGATCACC AGGTCGGCCT CCAGCTGCTC CTGCCGACCG

76051 CTGTCGAGGT CGCGGACGAC GACACCGGTG ACCCGGCCGC CACTGCCACC

76101 ACCACTTCCC GTCAGCTCGA CGGCCTCGGT GCGCTGCCGG ACGGTGATGT

76151 TGTCGGCTCC CAAGGCCTGC TGACGTACCG TCAAGTCCAG CAGCGGGCGG

76201 CTGGCGACCA GCGCGAACTG GGTGGCGGGG AAGCGGTGCT GCCACCCCTG

76251 ACCGGTCAGC GTCACCAGGT CCTCGGGGAA GCCGAGGCGG CGGGCGCCGG

76301 CCGCGAGGAG GCGGTCGGTG GTGCCGGGCA GCATCTCCTC GATGAGGCGG

76351 GCGCCGTTGG ACCACAGGAG GTGCGCGTGG CGGGCCTGCG GGACCCCCTT

76401 GCGGTGCTGG GGCTCCTCGG GCAGCGCGTC ACGTTCCACG ACGGTGACGG

76451 CGTCGACGTG CCGGGCCAGG ACGTGGGCCG CCAGGGTGCC TGCCATGCTG

76501 GCACCCAGGA CGACGGCATG TGCGGGTCGG GTGGTGGTCA CGCGCGTATC

76551 CCTTCGGGGT GGGTGGTGTC GGCGGGCCCG GCCGGATCGT CCATGGTCAC

76601 GTCCGTGACG CCCCAGAACG CCTGGACCCG GCGGCCGAGC CCGTGCTCGT

```
76651  CGAGTTCGAC GATGCCGACG ATGCGGAAGG TCATCGGCCG CGGCCGCTGC
76701  ACGGTGACCG TGGTCGGCGT CACCACGAAA CGGTCGTCCA TCGACGTCAT
76751  CGGCGGGTCC GGCACCTCGT GCGTACCGCA GGAGACGGCC AGTTCGAGAT
76801  GGCGGCGGAG ATCGTCCTTG CCCACCATCG GGGGCCGCCC CACGGGGTCC
76851  TCGAAGACGA TGTCGTCCGT GAACAGGTCG AGGACGCCTT CGATGTCACC
76901  GGCGTTGATG CGCTCGGCGT AGTCGACGGC CATCTGCTTG CGCGCGGCCT
76951  CGTCGGGCAT GGCACCTCCA GGAAGGGTGG GCAGACCTTG TGAAAGTCAT
77001  CGAGGGCCGT TCGGTTCAGC CGAGGACCGT GAGATCGGAT GTGCCCCAGT
77051  ACGACTTCAG ATGCCGGATG AGGCCGGACG CGTCCATGCG GATCACGAGC
77101  ATCGCCGTGC GGTGTATGCG GGCCGTCCCC GGGGCGTCGG GGGCCTTGAG
77151  CCAGCCCCGC TCCGCGTAGA GCGGGCCCAC GGGCAGGTAG TCCATGACGG
77201  AGGAAATCTG GATCAGCGCG TGCGTGGCGT CCTGCCCGGC GACGGGCTCG
77251  GCCGCCTCCT CGCGCAGGTG CGCGGCGAGC AGCGGTTCGT AGTGGGCGCG
77301  CAGCGCGTCG TGCCCGGTGA CGGGCGGGAG GCCGACCGGG TCCTCGAGGA
77351  CCGCGTCGGG CGCGTACAGA TCGATGATCG CGTCCAGGTC CCCGGCGTTG
77401  ATCCGCCGGC TGTGCTCCAG GGCCCGCTTC TTGCGGGCGA ACTCGTTCAT
77451  CGCTGCCCCT CCACTGCCTG ACCGTGTCCG TTGCCGTTGC CGTTGCCGTT
77501  GCCGTTGCCG TGTCCGTTGC CCTGCCCGGT GGGCTGTCCG TTGCCCTGTC
77551  CGCTCGCGCC GTCCCTGCCG AGGTCCCGGT CGATGAACGC GAAGATCTCG
77601  TCCGCCGACG CGTCCTGGAT ACGTGTACGA GTGGCCACCG CGACCTCGCC
77651  GGCCGTGTCC TGCGGCGCGT CGAGCCTGGC CAGCGTCGCG CGCAGCCGCC
77701  CCGCCAGTTC GGCCCGCGCC GAGCCGTCCT TCGAGGAGAC CGAGAGCAGC
77751  GAGTCCTCGA TGCGCTCGAA CTCCGCCAGG ACGTCGGCGA GCGGATCCGC
77801  CGCGCGCGGG GCCAGCTCCT GCCGCAGCTG CGCGGCGAGC TCCGCCGGGT
77851  TGGGATGGTC GAAGACGAAC GTGGCGGGCA GCTTCAGCCC CGTCGCGGCC
77901  GAGAGCCGGT TGCGCAGCTC CACCGCGGTC AGGGAGTCGA AGCCGAGTTC
```

77951 CCGCAGCCCC TGCGTCGCGT TGACGGGCGT GGCCGCGTCG TAGCCGAGGA

78001 CGGCCGCGAT ATGGGTGCAC ACCAGGTCGA GCAGCGCCTC CTCCCGCTCG

78051 GGGTCGGACA TCGCGCCGAG CGACTTGAGC AGCGCGGCCG CCCCCGCCGA

78101 CACGGCACCG CCGCCGCTCT TGCTCCCCCC GCGCACCAGG TCGCGCAGCA

78151 GCGCCGGTGC GGGGTGGCTC TGGGCCTGCC GGCGCATCCG GGCCAGGTCC

78201 AGACGGACCG GCGCGTACAG GGGCAGTCCG CCGGCCCACG CCGCGTCGAG

78251 GAGGGCGAGT CCCTCGTCGG CGCCGAGCCC GACCACGCCG GCGCGGGCAT

78301 GGCGCGCCCG GTCGGCGTCG GTGAGCCGTC CCGACATGCC GCTCGCCAGC

78351 TCCCAGTAGC CCCACGCCAG GGAGGTCGCC GCCGCACCGC CGTCGTGCCG

78401 GTGCCGGGCC AGCGCGTCCA AGAAGGCGTT GGCGGCCGTG TAGCTGCCCT

78451 GGCCGGGGCC GCCGAGCAGC CCGGCGACCG AGGAGTACAG GACGAACGCG

78501 GACAGGTCCG CGTCCCGCGT CAGCTCGTGC AGGTGCCACG CGGCGTCCGC

78551 CTTCACGCGC ATCACCTCCT CGACCTGCTC GGCCGTGAGG TTCTGCACCA

78601 CGGCGTCGTT CACGGTGCCC GCGCAGTGGA AGACGGCGGT CAGCGGGTGG

78651 TCCGAGGGCA CCGCCGCGAG GAGGGCGGCG GCTTCGTCCC GGTCGCCCGG

78701 GTCGCACGCG GCGAAGGTGA CTCGCGCGCC GAGCGCGGAG AGGTCGGCGG

78751 CCAGTTCGAG TGCGCCCGGC GCGTCGGCTC CCCGCCTGCT GGACAGCAAC

78801 AGGTGCCTGG CTCCGTACCG TTCCACCAGG TGACGGGCCG TCAGCGAGCC

78851 GAGTGCTCCG GTGCCGCCGG TGACCAGCAC GGTGCCCTCG GGGTCGAAGG

78901 CGGGAGGCAG CGAGAACACG GTCGTGCCCG CCGAGGGCGG GGCCGCCATC

78951 GCGGCGGGCG CCTGCCGGAT GTCCCACACG GTGATGTCGA GCGGCGTCAG

79001 AGCACGGCTG TCACCCCGTT CCGCGGGCAG CCCGGGCTCC GCCGACTCCG

79051 TGATCTCGGC AAGCTCGGTC AGCTCGGTCA GCTCCGCGAG GATTTCCCGT

79101 ACGCGCCCGG GCTCGGGCGG CACGACAGCC TGTCCCTCGT CCGGACGACC

79151 GCCCGCACGG TGGACCACCA GGGCCCCCTC GTGGCGGAGG GTGACGTCGG

79201 CCGCGCGCTC GGCGCCCGAA TCATCCGCCG TCGACGCACC GTCCACGGCC

79251 TCGACCCGCC ACCGCCCGGC AAGAGCAAGA CGCAGCACGG CACGGCCGAC

79301 GGAACCGGTT TCCTCCCCGA CGAGCAGAGT CTCGCCGCCG CGCGGCGCCA

79351 CGACATCCGC CAGCACGTGA TACGCGGACA CATAGGCCCC CAAGGACCCG

79401 GCCGCCTGCG CCCAACTCCA GCCCGCCGGA ACCGGCATAA GCAGCGCGGC

79451 ATCGGTGACG GCCACCGGGC CCACCGCGTC GAACAACCCC ATCACCCGGT

79501 CGCCCACGGC CACCGAACCG ACCTCGCCGC CGACTTCCGT CACCACACCG

79551 GCACCCTCGA CCTGGCCCGC CGTGAGCGGC CCCGGCGCCG CGGCCCGCAC

79601 CGCCACCCGC ACCTCGTGCG GCTCCAGCGC CCGTCCGGCC TCGGGAGCGT

79651 CGACAAGGGA CAACTGCTGT CCGCCGCCCG CCTCTTGGCA CCGAGCCAGC

79701 CGCCACGTGA GCGATCCGAC CGGCGGCACC AGCCGCACCG ACGCGTCGTC

79751 GCGCACGAGC CGTGGCACGT AGGCGCGCCC GTCACGCAGC GCCAATTCCG

79801 GTTCGCCGGA GGCCAGTACG CCGGTCAGCG TGGCCGGAGA AGACTCCAGT

79851 CCGTCCACGT CGAGCAGCGT GAGGCGACCG GGATTCTCGG CCTGCGCGCT

79901 GCGCACCAGA CCCCACAGCG ACGCGCCCGC CAGATCACCG GCGGTCTCAC

79951 CCGGCCGCGC GGCGACCGCG CCTCGGGTGA CGACGACGAG ACGGGTCGCC

80001 GCGAACGCCG GGTCGTCCAC CCACTCCTTG AGCAGCGACA GAAGGGACAC

80051 GGTGGCCAGC CGCGCGTACC CGGCCGGGTC GCCGCCCCTG CCATCGGCAT

80101 CCGCAACGGC CCCGGCACCT GCGCCGGGCG CGGCGCACAC GGCGAGCACG

80151 ACATCGGGCG CTTCGCCCCC AGCCGCCACT CCGTCCCGGA GCGCACCGAA

80201 CGTGTCCCAC ACGGGGCCGG CGGCCAGCGC ATCGGACAAG GCGTCGGCCA

80251 GCGCACCGGC CGACGTACCG CCCATCGGGC CACTCTCGAC CGGCGCGAGG

80301 ACCGCGGCAC GCGGGGCGCC GCCGCCCGTC TCCTCGGCCC GCGCGGCGAC

80351 CTCCATCCAC ACGAGCCGGA ACAGCGCGTC ACGGTCCGCC GCACGGGCGC

80401 CCGCGATCTG GTGGGCGGCC ACCGGCCGTA CCGTGAGCGA CTCCAGCGTG

80451 AGAACCGGCT CCCCGCCTCC GCCCCCGTCC ACGGCCGTGA GGGCCAGCTG

80501 GTCGGGCGCG GTGCGTGCGA TACGTACCCG CAACTTCTCA GCGCCCGGCG

```
80551  CGTGCACCCG CAACCCGCTC CAGGAGAACG GCAGCAGCAC TTGGTCGGTG
80601  TCGGCGGACG ACGTGACCGC GTCCAGGATC AGCGCGTGCA GCGTGGCGTC
80651  GAGCAACACC GGGTGCACCT GGTAGCGGTC GGCCCTGCCG CTCTCCGCCT
80701  CGGGCAGCGC CACCTCGGCG AAAAGGTCGT CCCCGAGCCG CCACGCGCTC
80751  ACCAGTCCCT GTGAGCCGGG CCCGAAGTCA TAGCCGTACG AAGCGAGTTC
80801  CCCGTACGGA TCCTGCTCGC CGACCGGTGT GGCGCCCGGG GGCGGCCACG
80851  TCCCGCCGAA CGAGGCGTCC CCGGCGTCGG GCCCCGGGGG AGCGACCACG
80901  CCCGCGGCAT GCCGGGTCCA CACGGCCTCC TCGCCCTCAC CCGTGGGCCG
80951  CGAATGGACG GTCACGGGAC GCCGCCCGTC CTCGGCCACG GAACCGACCA
81001  CCACCTGCAC GTCGACCGCG CCCGCACCCT CGTCCCCGAA GGCGAGCGGA
81051  GTGTGCAGCG TCAGCTCCGC CAACTCCGCG CAGCCGGCCC GCACCGCGGC
81101  CTGCAGCGCG AGCTCCACGA ACGCCGAACC GGGCAGCAGC ACCGTGTCCA
81151  TGACCCGGTG CTCGGCCAGC CACGCCTGGT CCCGCGGAGA GATCCGGCCG
81201  GTCAGCAGGT GACTGCCGCC GTCCGCGAGT TCCACGGCGG CTCCGAGCAG
81251  CGGATGCCCC GCGGACGCGA GCCCGAGCCC CGCCGGGTCC CCGGCGAGCC
81301  CCCTGCGCCC CTCCAGCCAG AACCGCTCCC GCTGGAAGGC GTACGTCGGC
81351  AGATCCACCA CCCGAGGCAG CGGCACGGCC GGGAACCAGC CCGTCCAGTC
81401  GACCTCCGCC CCCGCGCCGA AGGCCTGGGC GGCCGCGCGG GTGAGCTGCG
81451  CGGCGTCGCC GTGGTCGCGG CGCAGGGTGG GCACGACGGT GGCGGGCATG
81501  TCGGCCCGCT CGATGGTCTC CTCCATGCCG AGGTTGAGGA CGGGGTGGGG
81551  GCTGGCCTCG ATGAACAGGC GGTAGCCGTC GGCCAGCAGC GCTTCGATGG
81601  TGTCGGCGAA GCGGACGGGC TGGCGGAGGT TGGTGACCCA GTAATCCGTG
81651  TCGAGGGTGG TGGTGTCGTC GAGGCGTTCG GCGGTGACCG TGGAGTAGAA
81701  GGCGACGTCC GTGGTCGTGG GCCGGATGTC GGCCAGGCGC TCGGTGAGGA
81751  GGTCGTGGAG CTGGTCGATC TGGGGGCCGT GGGAGGCGTA TCCGACGTCG
81801  ATGACGCGGG CGCGCAGGCC TCGCGCCTCC GCATCCGCGA CCACGGCTGC
```

```
81851  CACATGCTCC GGCGGCCCTG AAATGACCGT AGAGGAGGGC CCGTTGACGG
81901  CAGCGACACA CACGCCGGGC CGGTCGCCGA TGAGCTCAGC AACCTGCTCC
81951  GAGCCGGCCC CCAACGACGC CATGTCGCCC TGCCCCATGA GCTGACGGAG
82001  CGCGTCACTG CGTACGGCCA CGATCCGCGC CGCATCCTCC AGTGACAGTG
82051  CCCCCGCCAC ACACGCGGCG GCCATCTCGC CCTGCGAGTG CCCGATGACG
82101  GCAGCCGGGC TGATGCCGTA ATCGGCCCAC ACCGAAGCCA GCGAGACCAT
82151  CACCGCCCAC AACACGGGCT GCACGACCTC GACCCGGGAC AGCTCACTCC
82201  CGTCCCCGCG CAACACCGCA CTCAGCGACC AGTCCACATG CGCCGACAGG
82251  GCCCGCTCAC ACTCCGCGAT CCGCGCCGCG AAGACGGGGG ACTCGTCAAG
82301  GAGCTGGGCA CCCATGCCCA CCCACTGCGA CCCCTGCCCC GGAAACACCA
82351  ACACCGGACC CGCGCCGGAG GCGCCCTGTA CGGCGCCCTC GACGACGTCC
82401  GGTGACGGCT CGCCCGCCGC CAGGGACCGT AGCCCGGCGA GGAGAGTCTG
82451  GCGGTCCTTG CCCACGACGA CGGCTCGGTT CTCGAACACC GACCGGGTCT
82501  TGACCAGGGA CCAGCCCACG TCCAGCGGCG ACGCGAGCCG CGGGTCGGCG
82551  GTGGCGCGGT CGGCCAGCAG GCGGGCCTGG GCCCGCAGCG CCTCCTCGCC
82601  GCGCGCCGAC ACCACCCAGG GCACCACTCC GGCCGGCGCC GCGGCGTCCT
82651  CCGCCGGAGC GGTCACGGGC TCCGGCGCGT CCGGGGCCTG TTCCAGGATG
82701  AGGTGCGCGT TGGTGCCGGA GATGCCGAAG GCGGACACCC CGGCGCGGCG
82751  CGGGCGTTCG CCGCGCGGCC AGGAGACCGG TTCGGACAGC AGGCGGACGC
82801  CACTGCCGTC CCAGTCCACG TGCGGCGTGG GCGCGTCGAT GTGCAGGGAG
82851  GCGGGCAGCT GTTCGTTGCG CAGCGCCATG ACCATCTTGA TCACACCGGC
82901  GACACCGGCC GACGCCTGCG CGTGCCCGAT GTTCGACTTG ATCGAGCCGA
82951  GCCACAGCGG CCGGTCCGCG GGCCGCTCCT TGCCGTAGGT GGCGACGAGC
83001  GCGCTGGCTT CGATGGGGTC GCCCAGCATG GTGCCGGTGC CGTGCGCCTC
83051  CACCGCGTCG ACGTCCTCGG CGGAGAGCCG CGCGTTGGCG AGTGCCTGCC
83101  GGATCACCCG CTGCTGCGCC TGCCCGTTGG GTGCCGTGAG CCCGTTGCTC
```

```
83151  GTGCCGTCCT GGTTGATGGC CGAACCCCGG ATCACCGCCA GGACGTTGTG
83201  GCCGTTGCGC CGGGCCTCCG AGAGCCGTTC GAGTACGACC AGGCCGACTC
83251  CCTCGGCCCA GCCGGTGCCG TCGGCGGCGG CCGCGAACGG CTTGCACCGG
83301  CCGTCCTTGG CGAGCCCGCG CTGCAGCGAG AACTCGACGA ACGAGCCCGG
83351  CGTGGCCATC ACCGTCGCGC CGCCCGCGAG AGCGAGCGAG CACTCGCCCT
83401  GGCGCAGCGC GTGCGCCGCC TGGTGGATCG CCACCAGGGA CGAAGAGCAG
83451  CCGGTGTCGA TCGTCATGGC GGGGCCTTCT AGGCCGAGTA CGTACGACAC
83501  CCTGCCGGAG GCGACACAGC CGAGGTTGCC GGTGCCGATG TAGCCCTCGA
83551  CCTCGGTGGG CTGTTCACCG ACGAGCGCGA GGTAGTCGAA GATGGTCAGG
83601  CCCGTGAACA CCCCGGCGTC GCTGCCCTTG AGGGTCTCCC GGTCGAGGCC
83651  CGCGCGTTCG ATCGCCTCCC ACGCGGTCTC CAGGAGCAGC CGCTGCTGCG
83701  GGTCCATCGC GACGGCCTCG CGGGGGCTGA TGCCGAAGAA TCCGGCGTCG
83751  AAGTCGCCCG CGTCGTAGAG GAACCCGCCT TCGCGCACAT AGCTGGTGCC
83801  GCGGCTCTCC GGGTCCGGGT CGTACAGCGT CTCCAGGTCC CAGCCCCGGT
83851  CGTCGGGGAA GGCCCCCATG GCGTCCTTGC CGGCCGCGAC CAGATCCCAC
83901  AGCTCCTCGG CGGAGCGGAC GTCGCCCGGA TAGCGGCAGG CCATGCCGAC
83951  GATCGCGATC GGCTCGTCGT CGGCGGCGCC CCTGGAGGCC CCGGCCGCCC
84001  GCACCGGGTC GGCGGAGGCC GCCGCGTCAC CGGACAGCTC GGCCCGCAGG
84051  ACGTCGGTGA GCGCGTCGGG GGTGGGGTGG TCGAAGACGA CCGTGGTCGG
84101  CAGTGTCAGG CCGGTGCTCT TGTTCAGCCT GTTGCGCAGC TCCACCGCGG
84151  TCAGCGAGTC GAAGCCCAGC TCCTGGAACG GCTTGGTGGC GGGCACCGCG
84201  TCGACGTCCG AGTGCCCCAG CGTGGCCGCC GCCTGGGAGC GCACGTGCTG
84251  CAGCAGCAAC TGCCGCTGCT GCGCCGGCTT CGCCTCCGTC AGCTCCTGCT
84301  GGAGCGACGA TGCCTCCGTG GCGTCTTCCT GCTGTGCCGC GGGTGCGCTG
84351  GCCCGCCGGT TCTCGGGCAG ATCGGCGAGG AGCGGGCTGG GCCGCTGCGC
84401  GGTGAACGTC GACGTGAACT GCGCCCAGTC GAAGTTCGCC ACGGTCAGCG
```

84451 TCGTCTCACC CGCGTCCAGG GCCTGCTGCA GCGCCTTGAC GCACAGCTCC

84501 GGGCTGAGCG GGTGCAGGCC GAAGCGGCTG AAGAACGTCA ACGCGGCCTG

84551 GTCCGCCGCC ATGCCCGCCT CGGCCCAGGG CCCCCAGGCG ATGGAGGTGG

84601 CGGGCAGGCC CTCGGCGCGG CGGTGCTCGG CGAGGGCGTC GAGGAAGTGG

84651 TTGGCCGCAC CATAGGCGCC CTGCTGGCCA CTGCCCCACA CGCCTGCGCC

84701 CGACGAGAAC ATCACGAACG CCGAGAGCGG CAACTCCCGG GTCAGTTCAT

84751 GCAGATGGTG AGCGGCGAGC GCCTTCGGAC GCAGCACCTC GTCCAGCTCG

84801 GCACCCGACA CGTCGCCGAG ACCGATGTAG TTCGGCACGC CGGCCGCGTG

84851 GATGACGGCG GTCAGCGGGT GCTCGGCGGG GACATCGTCG ATGAGGCGTC

84901 GCACCTGCTC GCGGTCGCCG ACGTCGCAGG CGGTGACGGT GACGGCGGCC

84951 CCCAACTCCG TCAGTTCCGC GGCGAGTTCC TGTGCTCCCG GGGCGTCGGG

85001 GCCGCGGCGG CTGGTCAGGA GGAGGTGCGG GGCGCCCGCA CGGGCGAGCC

85051 ACCGCGCGAG GACGGCGCCG ATGCCGCCGG TCCCGCCGGT GATGAGAGTG

85101 GTGCCGTCGG GCCGCCAACC AAGCCCGCTG CCGACCGTGT TGGCGGGCGC

85151 GTGTGCAAGG CGACGGGCAT GGACGCCGGA CGGCCGGATG GAGATCTGGT

85201 CCTCGTCCTG CGGAACCAGC GCGGCGGCCA GCCGGGCCAG CGTCTGATGG

85251 TCGATACGAG CGGGCAGATC GACCAGCCCG CCCCACAGCC GCGGATACTC

85301 CAGCGCAGCG ACGCGCCCCA GCCCCCACAC CTGAGCCTGC ACCGGGTGGG

85351 TGAGGGCGTC GCCGGCGCTC GTGGAAACAG CCCCCTGCGT GAGAGTGCGT

85401 ACGGCGATGT CGGCGCCGTT GTCCGCGAGG GCCTGGACGA GAGCGGTCGT

85451 CGCGGCGAGT CCGGCGGGCA CGGCCGAGTG CTCGGGATGC GGCTCCTCGT

85501 CCAGGGCCAG CAGATTGACG ACTCCGGCAA ACGCGGCCCC GTCCATCAGG

85551 ACACGCAGCT CCTGCGCCAA CTCCGTACGC TCCATGGCAC GTGCGTCGAC

85601 CACGTGGCGT CGCACCTCGC CACCATGGGC GGTCAGCGTC TGCGCGGTCG

85651 CGAGGACGGC CGGGTGGTCG GCGTGCGCGG CGGGCACGAG CAGCAGCCAG

85701 GCCCCGCTGA GCTCCGGCGC CGGCACGTCG GGCAGATGCT TCCAAGTGAC

85751 CTGATAACGC CAGGAGTCGA CGGTGGACTG CTCGCGGTGC CGACGCCGCC

85801 AGGCCGAGAG GACGGGCAGC GCGGACTCCA GCGCTCCGAC GCTCTCCGCC

85851 TGCCCCTCGA TCTCCAGACT GCCGGCGAGG GCGTCGATGT CCAGGTCCTC

85901 GATCGCCTGC CACACCCGGG CCTCGACCGG ATCGTGCCCA CCACCCACGG

85951 CTGCGACCGC CGCGGGCGGC TCCACCCAGT AGTGCTTGTG CTGGAAGGCG

86001 TAGGTGGGGA GGTCGACGGT ACGGGGGGTG GGGTCGGCCG GGAACCAGCG

86051 CCGCCAGTCG ACGGGGGCGC CGGCGGTGAA GGCGTGGGCG GCCGCGCGGG

86101 TGAGCTGGGT GGTGTCACCG TGGTCGCGAC GCAGGGTGGG GATGGTGACG

86151 GCCGTCCCCG CAGCACCGGC CTGCTGCTCG ATGGTCTCCT GGATGCCGAG

86201 GTTGAGGACG GGGTGGGGGC TGGCCTCGAT GAACAGGCGG TAGCCGTCGG

86251 CCAGCAGCGC TTCGATGGTG TCGGCGAAGC GGACGGGCTG GCGGAGGTTG

86301 GTGACCCAGT AGGCGGTGTC TAGGGCGGTG GTGTCGTCGA GGCGCTCTGC

86351 GGTGACCGTC GAGTAGAACG CCACGTCGGT GGTGGTCGGC TGGATGTCGG

86401 CGAGCCGGTC GGTGAGGAGG TCGTGGAGCT GGTCGATCTG GGGACCGTGG

86451 GAGGCGTACC TGACGTCGAT GACGCGGGCC CTGAGTCCCT GCGCCTCCGC

86501 ATCGGCGACG ACGGCTGCCA CATGCTCCGG CGGGCCCGAA ATCACGGTCG

86551 ACGACGGTCC GTTGACGGCC GCGACGACTA CGCCCGGCCG GTCGCCGATC

86601 AGCTCTGCGG CCTGCTCGGC ACCGGTGCTG AGCGAGGCCA TGTCGCCGTG

86651 CCCTTGCAGC TGACGAAGCG CGTCGCTGCG TACGGCTACG ATCCGTGCCG

86701 CATCCTCCAG TGACAGTGCC CCCGCCACAC ACGCGGCAGC CATCTCGCCC

86751 TGCGAGTGCC CGATGACGGC AGCCGGGGTG ATGCCGTAAT CGGCCCACAC

86801 CGCAGCCAGC GAGACCATCA CCGCCCACAG CACGGGCTGC ACGACCTCGA

86851 CCCGGGACAG CTCGCTCCCG TCCCCGCGCA AGACATCACT CAGCGACCAG

86901 TCCACATGCG CCGACAGCGC CTGCTCACAC TCCGCGATCC GCGCCGCGAA

86951 GACGGGCGAC TCGTCAAGGA GCTGGGCGCC CATGCCCACC CACTGCGACC

87001 CCTGCCCCGG AAACACCAAC ACCGGCCCAG GACCCACATC ACCGGCCACC

87051 CCGGCCACCA CATCCGCCGA CGCCTCACCC GCGGCCAATG CCTCCAGGCT

87101 GGCACCAGCC TGAGCCAAGT CCCGCCCCAC GACCACGGCC CGCTGATCGA

87151 ACAACGCGCG TGTCGTGGCC AGCGACCAGC CCACCTCGGA GACCGACGCA

87201 TCCGCCAGCC CGGCCGCGAA CTCGCCCAGC CGCCGCGCCT GTTCACGCAA

87251 CGCGTCCGGC GTCCGCCCGG ACACCACCCA CGGCACGACC CCACCCGGCT

87301 CAGCCGCCAC GGGGCCCGGC GCGTCCTCTT CCGGCGGCGC CTCCTCCAGA

87351 ATCAGGTGCG CGTTCGTCCC GGAGATCCCG AACGCCGAGA TGCCTGCCCG

87401 CCGCGTGCGC TCCGCCGGCC AGTCCACGGG CTCGGAGAGC AGTCGTACGC

87451 TGCCCTGTTC CCACTGGACG TGCGGTGACG GGGCGTCGAT GTGCAGGGAG

87501 GTCGGGAGGA GACCGTTGCG CATCGCCATG ACCATCTTGA TGACGCCCGC

87551 GACACCGGCG GCGGCCTGCG TGTGGCCGAT GTTGGATTTC ACCGAGCCGA

87601 GCCAGAGCGG ACGGTCCTCC GGGCGCCCCT GGCCGTACGT GGCGATCAGG

87651 GCCTGCGCCT CGATGGGGTC GCCGAGCGTG GTGCCGGTGC CGTGCGCCTC

87701 TACGGCGTCG ATGTCCTCGG CGGAGAGGCG GGCGTTGGCG AGGGCGGCGC

87751 GGATGACGCG TTCCTGGGAG GGGCCGTTGG GGGCGGCGAG CCCGTTGCTC

87801 GTACCGTCCT GGTTGGTGGC CGAACCCCGT ATCACCGCAA GGACCTTGTG

87851 GCCGCGGCGC CGCGCTTCGG AGAGCAGCTC CAGCGCCACC ACCCCGGCGC

87901 CCTCGCCCCA GCCGGTGCCG TCGGCGGCGG CCGCGAACGG CTTGCACCGC

87951 CCGTCGGGCG CGAGCCCCCG CTGCCGGGAG AACTCGGTGA ACGAACCCGG

88001 CGTCGCCATC ACCGTCGAAC CGCCCGCCAG CGCGAGCGAG CACTCGCCCT

88051 GCCGCAGCGC CTGACTTGCC AGATGGATCG CCACCAGGGA CGACGAGCAC

88101 GCCGTGTCGA CGGTGACCGC GGGACCTTCG AGCCCCACCG TGTAGGAGAT

88151 CCGGCCCGAC ACCACACTGC CGAGGTTGCC GGTGCCGATG TACCCCTCGA

88201 CGTCGCTGGC CGTCTGGCTG ATCAGCGTCA GGTAGTCGTG GGCGCTCACT

88251 CCGGTGAAGA CGCCGGTGTC GCTGCCCTTC AGCGCGTGCG GGTTCATGCC

88301 CGCGTGCTCG ATCGCCTCCC ACGCGGTCTC CAGGAGCAGC CGCTGCTGCG

88351 GATCCATCGC CGTGGCCTCG CGCGGGCTGA TGCCGAAGAA CTCGGCGTCG

88401 AAATGGCCGG CGTCGTACAG GAAGGCGCCG TCCCGCACAT AGCTGGTGGC

88451 CGGATGCTCC GGATCCGGGT GATACAGCGA CTCCAGGTCC CAGCCCCGGT

88501 CGTCGGGGAA CCCCGCGACC GCGTCACCCC CGTCGCGCAC GAGTTCCCAG

88551 AGGTCCTCCG CCGACCGGGC GCCGCCCGGG TAGCGGCAGG CCATGCCGAC

88601 GATGGCGACC GGCTCGGTCG ATTCCTTGTC GTGGAGCCGT TGCCGGGCCT

88651 GGCGCAGCTC CGCGGTGACC CACTTGAGGT GATCGAGAAG CTTCTCCTCG

88701 TTCGACATCT GACCCAGGCT CCTTGGCGCT ACGTGGTGAT CGGGGCGTAT

88751 GAGGTTGGGG GAGGGCAAGG GGGCCGGTGT GGCCGGGGCT CATCGCGCTC

88801 AGGACTGATC GCTGCTCAGG ACTTCCCGAA CTCACTGGAG ATGAGGTCGA

88851 AGATGTCGTC CGCGCTCGCC GCCTCCAGAT CGGCATGGGC CGAATCAGTG

88901 CCTTCCGGCC CGTCCTGCGC CGGACTCCAC TTCGACACAA GGACCTGCAG

88951 CCGGCCCACG ATGCGGCGCC GGGCCGCCTC GTCCACCTCG GCCGCTCCGA

89001 ACGCCGTGTC CCACTTGTCG AGCGCCGCGA GCACGTCGCC CTCACCTGCG

89051 ACCTCGGCGC CGTCGCCGAG CTGTCCGCGC AAGTGCGTGG CGAGGGCCTC

89101 GGGCGTGGGA TGGTCGAAGA TCACGGTGGC GGGGAGCGAG AGTCCGGTCG

89151 TGGTGTTGAG CTGGTTGCGC AGCTGGACCG CGGTGAGCGA GTCGAAGCCC

89201 AGCTCCTGGA ACGGCTTCGC GGCGGGAATG TCCTCCACCG TGCGGCCGAG

89251 CGTCGCGGCC GCGTATGTCC GGACCTGCTG GACCAGGAAG CCGAGCCGCT

89301 GTGATGCGGG CGTCTTCGCC AGCTCCTCGC GGAACGCGCT CGTCTCGGCG

89351 GCGGTCCCCG TCTGCTCGGC CTCCCGCTGG TTCTCCGGAA GGTCGTCGAG

89401 GAACGGACTG GGCCGCTGCG CGGTGAACGT CGGCGTGAAC TTCGCCCAGT

89451 CGAAGTTCGC CACGGTCAGC GTGGCGTCGC CCGCGTCGAC CGCCTGGTGC

89501 AGCGCCTTGA CGCACAGATC CGGAGCGATC GGGAGCAGAC CGAAGCGCTT

89551 GAAGTACGTC AGTGACTCCG GGTCGGCGGA CATGCCCGCC TCGGCCCAGG

89601 GCCCCCAGGC GATGGAGGTG GCGGGCAGGC CCTGGGCGCG GCGGTGCTCG

89651 GCGAGGGCGT CGAGGAAGTG GTTGGCCGCA CCATAGGCGC CCTGCTGGCC

89701 ACTGCCCCAC ACGCCGGCGC CCGACGAGAA CATCACGAAC GCCGAGAGGT

89751 CCAGGTCGCG CGTCAACTCG TGCAGGTTCC AGGCCGCGTC GGACTTCGAC

89801 CCCAGCACCT CGCCGAGGCG CGCGGTCGTC AGATCACCGA TCGCGGTCAG

89851 ATCGGTCATG CCCGCCGCGT GGATGACGGC TGTGAGGGGA TGCTCGGCCG

89901 GCATGTCGTC GATGAGGCCG CTCAGTTGGC GGGGATCGCT GACGTCGCAG

89951 GCGGTGATGG TGACGGCGGT GCCGAGCCCG TCGAGCTCGG CGGCGAGTTC

90001 CCGGGCGCCG GGCGCGTCGG GCCCGCGACG GCTGGTGAGG TGAAGACGGG

90051 GGGCGCCCTG CCGGGCCAGC CAACGGGCGA GGACGGCACC GATGCCGCCG

90101 GTCCCGCCGG TGATCAGGGT GGTGCCCCGA GGCCGCCAGG TGGCCTCGCT

90151 GTGCACGGGA TTCTGAATGC TTCCGACGGC GTGCGTGAGG CGCCGGTGGT

90201 GGATTCCGGT GGGGCGGACG GCGGTCTGGT CCTCGTCGTC CTGGGGGAGG

90251 AGAGCGGCGG CGAGGCGGGG GAGGGTGTGG CGGTCGATAC GAGCGGGGAG

90301 GTCGACGAGT CCGGCCCAGA GGCGCGGGTG TTCGAGGGCT GCGACGCGGC

90351 CGAGCCCCCA GACGTGAGCC TGGAGGGGGT GGGTGAGTGG GTCGGTGGCG

90401 CCCGTGGACA CGGCACCCTG CGTGACGGTG TGCAGGGGTG CGGTCGTGCC

90451 GTTGTCGCCG AGGGCCTGGA GGAGAGCGGT CGTCGCGGCG AGCCCCGCGG

90501 GCACGGCGGG GTGCTCGGGG TGCGGCTCCT CGTCCAGCGC CAGCAGATTG

90551 ACGATTCCGG CAAGACCGGC CGTGTCCACC GCGGCCAGCT CCTGACGTCC

90601 CGCCCGGCCG GTCTCGACCG GATGCAGCCG GACGGCGGCC GCCCCGTGCT

90651 CGCTCAACGC CTCGGCGGTG GCTCGTACGG CGGGGTGCTC CGCCTTGTCG

90701 GCAGGGACGA ACAGCAGCCA GTCGCCGCCG AGTTCCGGTG CGGGCCCGTC

90751 GGACCGCTGT TTCCACGTGA CGCGGTACCG CCAGGAGTCG ATGGTCGCCT

90801 GGTCCTGGTG CCGACGCCGC CAGCCCTTGA GCACCGGCAA CGCGGGCTCC

90851 AGCGCCCGGA CCGCCTCCTC GCTGCCCTCC TCCGACCCCA GCGTCTCGGC

90901 CAGCAGACCG AGATCGAGCT CCTCGACGGC GTGCCACAGC TGGGCCTCGG

90951 CCGCACTCTG CTCACCGCTG ACGGCGCCCG AGGCGGACGC GGAACGTTCG

91001 AGCCAGTAGT GCTGGTGTTG GAAGGCGTAG GTGGGGAGGT CGACGGTGCG

91051 GGGGGTGGGG TCGGCCGGGA ACCAGCGCCG CCAGTCGACG GGGGCGCCGG

91101 CGGTGAAGGC GTGGGCGGCG GCACGGGTGA GCTGGGTGGT GTCGCCGTGG

91151 TCGCGGCGGA GGGTGGGGAC GACGGTGGCG GGGATGTCCG CCTGCTCGAT

91201 GGTCTCCTCC ATGCCCAGGC CCAGCACGGG GTGGGCGCTG GCCTCGATGA

91251 ACAGGCGGTA GCCGTCCGCG AGAAGGGCTT CGATGGTGTC GGCGAACCGG

91301 ACCGGCTGGC GGAGGTTGGT CACCCAGTAA TCCGTATCCA GGGCTGTGGT

91351 GTCCGTCAGA CGCTCGGCGG TGACCGTCGA ATAGAAGGCC ACGTCCGTGT

91401 TCGCGGGCCG GATGTCAGCC AGGCGTTCGG TCAGCAGATC GTGGAGCTGG

91451 TCGATCTGGG GGCCATGCGA GGCGTACCCG ACGTCGATGA CACGGGCGCG

91501 CAGACCACGT GCCTCCGCAT CGGCGACCAC GGCAGCCACA TGCTCCGGCG

91551 GCCCTGAAAT CACCGTAGAC GACGGCCCAT TGACCGCCGC GACGACCACG

91601 CCCGGCCGGT CACCGATCAG CTCAGCGGCC TGCTCGGCAC CGGTGCTCAG

91651 CGAGGCCATG TCACCGTGCC CTTGCAGCCG ACGAAGCGCG TCACTGCGTA

91701 CGGCTACGAT GCGCGCCGCA TCCTCCAGCG ACAGCGCCCC CGCGACGCAC

91751 GCGGCAGCCA TCTCACCCTG CGAGTGCCCG ATCACAGCAG CCGGAGTGAC

91801 CCCGTAATCA GCCCACACCG CAGCCAGCGA GACCATCACC GCCCACAACA

91851 CCGGCTGCAC GACCTCGACC CGGGACAGCT CACTCCCATC CCCGCGCAAC

91901 ACCGCACTCA GCGACCAGTC CACATACGCC GACAGCGCCC GCTCACACTC

91951 CGCAATCCGC GCCGCGAAGA CGGGGGACTC GTCCAGCAGC TGGGCACCCA

92001 TGCCCACCCA CTGCGACCCC TGCCCCGGAA ACACCAACAC CGGCCCAGGA

92051 CCCACATCAC CAGCAACCCC GGCCACCACA CCCGCCGAAG CCTCACCCGC

92101 AGCCAACGCC CCCAGGCCAG CCGTCAACGC ATCGCGGTCA CGCCCCACCA

92151 CCACAGCCCG GTGCTCGAAC ACCGACCCGG TCGTGGTCAA CGACCAGCCC

92201 ACATCAGCCG CCGACGCATC CGCCGGCCCG GCCGCGAACT CGCCCAGCCG

92251 CCGCGCCTGT GCACGCAGCG CGTCCGGCGT CCGCCCGGAC ACCACCCACG

92301 GAACGACCCC ACCCGGCTCC TCGGCCACGG AGCCCGGCAC GTCCTCCTCC

92351 TCCGGTGGTG CCTCCTCCAG GATCAGATGC GCGTTCGTCC CCGAGAAGCC

92401 GAACGAGGAC ACCCCCGCCC GGCGCGGGCG CTCGCCCCGG GGCCACTTCA

92451 CCGGGTCGGT GAGCAGGCGC AGCCCGCTGC CGTCCCACTC CACGTGGGGC

92501 GAGGGGGCGT CGACGTGCAG GATGGCGGGC AGCAGGTCGT GCCGCAGGGC

92551 CAGGACCATC TTGATGACAC CGGCCACACC GGCGGCGATC TGCGTGTGGC

92601 CGATGTTGGA CTTCACCGCT CCCACCCACA GCGGCCGGTC CTCCGGCCGT

92651 TCCCGGCCGT AGGCGGAGAT GAGAGCCCCG GCCTCGATGG GGTCGCCGAG

92701 CGTGGTGCCG GTGCCGTGCG CCTCCACGGC GTCGATGTCC TCGGGGGCGA

92751 GGCGGGCGTT GGCGAGGGCG GCGCGGATGA CGCGTTCCTG GGCGGGGCCG

92801 TTGGGGGCGG TCAGGCCATT GCTCGCGCCG TCCTGGTTGA TCGCCGAACC

92851 CCGGATCACC GCGAGGACCT TGTGGCCCTT CCTGCGGGCG TCGGAGAGAC

92901 GCTCAAGGAG AACCACCCCC GTACCCTCCG CCATGCCCAT GCCGTCGCTG

92951 CTCGCCGAGA ACGGCTTGCA CCGTCCGTCG GGGGCCAGGC CGCGCAGTTC

93001 GCTGAAGCCG ATCAGCGGGG CGGGCGACGA CATCACGTAC GTGCCGCCCG

93051 CCAGCGCCAG CGAGCACTCC TGTGTGCGCA GGGCCTGGGT GGCGAGGTGA

93101 AGGGAGACCA GCGACGAGGA GCACGCCGTG TCGACCGTCA CCGCGGGGCC

93151 TTCGAGGCCC AGGGTGTAGG CGACGCGGCC GGAGGTGACG CTGCCGGAGT

93201 TGCCGATGGT GAAGTATCCG GCGGTGCCCT CGGGGACCTC GGACGCGCCG

93251 AGGGCGTAGT CGAGTCCGTC ACAGCCGATG AAGGTGCTGG TGTCGCTGGA

93301 GCGGAGGCTG AGGGGGTCGA TGCCGGCCCG TTCGATCGCC TCCCACGCCG

93351 TCTCCAGGGC GAGCCGCTGC TGCGGCGCCA TGGCCGCGGC CTCGGTGGGT

93401 CCGATGCCGA AGAAGGTGGG GTCGAAGTCA CCGGCGTCGT AGACGAAGCC

93451 GCCTTCCCGG ACGTAACTGG TGCCGGTGCT CTCGGGGTCC GGGTCGTAGA

93501 GGGAATCGAG GTCCCAGTTG CGGTTGCCGG GCAGGGGCGC GACGGCGTCG

93551 CCGCCGGTGG AGACCAGCTC CCAGAACTCT TCGGGAGACC GGACTCCGCC

93601 GGGCAGCCGG CAGGCCATGC CGATGACCGC GACCGGTTCG TGGCCCGCCG

93651 ACTCGACGTC CTGCAGCCGG CGTTCCGTCT GACGCAGGTC CGCGGTGACA

93701 CGCTTGAGGT ATTCCAGAAG TTTCTCTTCG GTGTGCGCCA TCCCGGTGAC

93751 AACCGCCCCT CTCCGCGAGA ACAGACCGCA GACTCGTCGA CGGCGCTAAA

93801 GCCCTCCTAA TACTCGGCTG TGTACCGCTC GCTGCCACGG GTGTCCGCAC

93851 TGGTCGGAGG CTCCGGCCCA GGGAACAGGG GCTTTCTTAG GGGCGCTTAA

93901 GCGGTGCCTG CCAGGGTGTG CCGGTGTCAG GCCGTCACGC CCTGATCAGC

93951 GGCGTCGCCC GTGCCGTGCC CGTGCGGTCG GTGGGCCTGA CCGTCGGTCC

94001 GGACAACGCG AAGCGAGGCA TCGTGCCCAT CACGGATAGC AAGCCGGCCG

94051 CCACATTCCC CGACCTGGTC GACCCGTCGT TCTGGGCGCG GCCGCACGCG

94101 GAACGCGTGG CGCTGTTCGA GGAGATGCGC GGGCTGCCGC GGCCGGCGTT

94151 CATCCGGCAG AACATGCCCG GCGTGCCCTG GACGTTCGGC TACCACGCGC

94201 TGGTCAAGTA CGCGGACATC GTGGAGGTGA GCCGCCGCCC GCAGGACTTC

94251 TCCTCGAACG GCGCGACCAC CATCATCGGT CTGCCGCCCG AGCTGGACGA

94301 GTACTACGGC TCGATGATCA ACATGGACAA CCCGGAACAC TCGCGGCTGC

94351 GGCGCATCGT CTCGCGTTCC TTCGGCCGCA ACATGATCCC CGAGTTCGAG

94401 GCCGTGGCGA CCCGCACCGC CCGCCGCATC ATCGACGAGC TCATCGCGCG

94451 GGGACCCGGC GACTTCATCA GGCCCGTCGC CGCGGAGATG CCCATCGCCG

94501 TGCTCAGCGA CATGATGGGC ATCCCGGCGG AGGACCACGA CTTCCTCTTC

94551 GACCGGTCCA ACACGATCGT CGGCCCCCTC GACCCGGACT ACGTGCCGGA

94601 CCGGGCGGAC TCCGAACGCG CGGTGATCGA GGCGTCACGC GAACTCGGCG

94651 ACTACATCGC TGGCCTTCGT GCGGAACGGC TCGCCGCCCC CGGCAACGAC

94701 CTCATCACCA AGCTCGTGCA AGTCCAGGCG GACGGCGAGC AGTTGACGCG

94751 GCAGGAACTC GTCTCCTTCT TCATCCTGCT CGTCATCGCC GGGATGGAGA

94801 CCACCCGCAA CGCCATCTCG CACGCGCTGG TACTGCTGAC CGAGCATCCC

```
94851  GAGCAGAAGC AGCTGCTGCT CTCGGACTTC GACACGCACG CGCCGAACGC
94901  GGTCGAGGAG ATCCTCAGGG TCTCCACGCC CATCAACTGG ATGCGGCGCG
94951  TCGCCACCCG CGACTGCGAC ATGAACGGCC ACAGGTTCCG CAGGGGCGAC
95001  CGGATCTTCC TGTTCTACTG GTCGGGCAAC CGGGACGAAT CCGTCTTCCC
95051  TGACCCGTAC CGGTTCGACA TCACGCGCGG GACGAACGCG CACGTCACGT
95101  TCGGCGCGGT GGGCCCGCAC GTCTGCCTCG GGGCCCACCT CGCCCGTATG
95151  GAGATCACCG TCCTGTACCG GGAGCTGCTC GCGGCGCTGC CCCAGATCCA
95201  TGCCGTGGGG CAGCCCCGCA GGCTGGACTC CAGCTTCATC GAAGGGATCA
95251  AGCACCTGCA CTGCGCCTTC TGAGCACATA CGCTTCCCTC TGCGCATGTG
95301  CGCTCACGAC GCTCCGATCA GCGACTGCCA ACGACTGTCA GCGACCGGAC
95351  AGGGCCAAGG GCGGTGGGGA CATCAGGTGC ATGTCACCCG CGAGTATGGC
95401  CCGCTGCAGC TCCTGGAGCG GGCGCCCGGG TTCGAGCCCC AGCTCGTCGT
95451  TGAGCGTCTT GCGCACCGAC TGGTACACCT TCAGCGCGTC CGCCTGCCGC
95501  TCGGAGCGGT AGAGCGCCAG CATCAGCTGG CGGTAGAACG CCTCGCACAT
95551  CGGGTTCTCC GCGGTGAGGG CGTACAGCAT GCCCACGGCC TCGCGGTGCC
95601  GGCCGAGCTG GAGCTGGCAC TCGACGAGCA TCTCCTGACA CTCCAGGCGG
95651  ATCTCGGTCA GCCAGGTCGA GAAGCCGTCG ATGATCGGGC CGTTGGTGCC
95701  GGGACCGTTC CCGCCCTGCC CGAGGATCGG GCCGCGCCAC AGCGCGAGCG
95751  CCTGCCCGAA ACAGGAGGCC GCCTCGTCGA ACCGCTTCTC CCTGAGCAAC
95801  GACCGCCCCA CGTCCACCAG TTCGGGGAAG ATCTGGGCAT CGATCTGGTC
95851  GTCGTCCCGC TTGTGCAGGA CGTACCCCGG CGCACGGGTC TCGACGGGGT
95901  TGCCCGCCGA ACCGGGCACC TTGAGGAACT TGCGGAGCTG GGAGATGTAC
95951  ACATGCAGTC CCGCCGTGGC GCGCCGCGGC AGGTCCTCGC CCCAGATCTC
96001  CCGCATCAGC TGCTCCAGGG AGACCACCCG GTCGGCGCGG ATGAGGAGCA
96051  CGGTGAGGAC GATCTCCACC TTCTGGGCGT TGATGGTGGC GTAGTCGTTT
96101  CCGTCCTTGA TGCGGAGCGG GCCCAGCATT TCGTATCTCA CCGAGCGTTC
```

96151 CCCCTTGCTG TCGCACGCTG CTGCGCACTG TCGGCCAGGG CCTTGGAGAT

96201 GACTTCCGTG ACGCCCTGCT GGTGCGTGTT CAGATAGAAG TGGCCCCCGG

96251 CGAAGACCTT GAGGTCGAAG GGGCCTTCCG TGTGCTGCTG CCAGGCCTCG

96301 ACCTCGTCCA GCGGCGCCTG CGGGTCCCGG TCGCCCACCA GGGCGGTGAT

96351 GGGGCAGGAC AGCGGCGGCG ACGGGTTCCA CCGGTACAGC TCGACCGCCC

96401 GGTAGTCGTT GCGGACGACC GGGATGATCT CCGCGAGCAG TTCCTCGTCG

96451 TCCAGGAACC GCGGGTCAGT GCCACCGGCC CGGCGCAGCT CGGCGGCCAA

96501 CTCGGTGTCG TCGAGGAGGT GTACGGTGCC GCGCCGGAAG CGGGACGGCG

96551 CGCGGCGTCC CGAGACGAAC AGCCGGCAGG GCTGCTTCCC CGTGCGCTCG

96601 CGGAGCCGCT GGGCGACTTC GTAGGCGAGG ACGGCGCCCA TGCTGTGGCC

96651 GAAGAACGCC AACGGGCGGT CGTCGAACGG GCCGAGCGCA TCGGTGATGA

96701 GGTCGGCGAG TTCCCCGATG TCGTCCAGGA GCCGCTCTCT GCGGCGGTCC

96751 TGTCGCCCGG GGTACTGCAC CGCGAGGACC TCGCTGTCGG TCGGGAGAGT

96801 GGGGGATTGC GCAAGGGGGT GGTAGTAGGA GGCCGAGCCG CCCGCGTGGG

96851 GGAAGCAGAC CAGGCGAACG ACGGCTTCCG GTCGGGGCCG GAAGCGACGT

96901 ATCCAAGGGT CCGACATATC GGGTGGGGGG AAGGCAGACA AGATCTTTCC

96951 CTTCGCCAGG AACGCTGACA ACGGTGTGTC GCCACATCAC ATAGCCGCTC

97001 CTGATCATGC GCAGCTCAAA GTTTAAACGG CAACGTCGCT AACGGGGGAG

97051 CAGGGCGGAA TCAGACATTC CCCATCCTTT ATTCCGCGAT TCTTACGTGA

97101 TCGAATCCCG GCGGCCAAGA TGGAGTAAAT TTCAATATGA ATGCTTAACG

97151 CCGCACAGCT TGTACGGCGG GCCGCCCGGG CGGTGACTGG CGTCCCTGCC

97201 AGCCGTGATG GCCTGACGAG GCCTCCGGGA TCCATCCCCC GCCCGCTGTC

97251 GCCGAGTTCT TTGCGGGATT ATTACGTTGC ATTGGTTTGC TTCGTGGCCC

97301 GGGCCGTTGG CCTGCGCTAT TTGGCAGCCT TCCGTCATGG GTGGTAAAAG

97351 ATCGCCTTTC CCCTCTGGGG TGCCGGTCGA GCTGGCCTCG ACCGCGATTG

97401 TGGCTTGTTG TTTTCTTGTG GCGCCGCGTG TGAAACAGCG GCAGTTGGCC

97451 ACTCGCTCTG ACAGGCTCCG GGGACGGGGT TGTCACCTTT TGGGGTGACT

97501 GGCCTCGTTC AAGGCGTCCT GGCCCGTGGT GCATCCGCGA TCGTCGTGCC

97551 ATGGGTGAAG TGGGAAGGAG CACAGAACGA TGAGCGAGAG CATGGCGTGG

97601 CTGACGCGGG ACGTCCGCAA GGCCCGCAAG GAGGGCAGTG CGGGGACCGC

97651 GCGGCGCCGA GCCGACCGGC TGGCGGACCT GGTCGCCCAC GCCCGCTCGG

97701 CGTCGCCGTA CTACCGGGAG CTCTACCACG GCCTGCCCGA GCGGATCGAG

97751 GACCCGACGC TGCTGCCGGT GACGGACAAG AAGCAGCTGA TGGACCACTT

97801 CGACGACTGG CCGACGGACC GCGACATCAC CTTCGAGAAG GTCCGCGCGT

97851 TCACCGACGA CCCCGAGCTG ATCGGGCGGC GCTTCCTCGG CCGCTATCTG

97901 GTGGCCACCA CGTCGGGCAC CAGCGGCAGG CGCGGCCTGT TCGTGCTCGA

97951 CGACCGGTAC ATGAACGTGT CCTCCGCCGT CTCCTCCCGG GTGCTCGCCT

98001 CCTGGCTCGG CCCCCTCGGC ATCGCCCGGG CCGTCGTCCA CGGCGGCCGC

98051 TTCGCCCAAC TCGTCGCCAC CGAGGGACAT TACGTCGGCT TCGCCGGATA

98101 CTCCCGCCTG CGCCAGGACG GCGGAGCGCG CAGCAAGCTC GTCCGCGCCT

98151 TCTCTGTGCA CGAGCCGATG TCACGTCTGG TCGCCGAACT CAACGAGTAC

98201 CGGCCCGCGT TCGTCATCGG CTACGCCAGT ACGATCATGC TGCTCACCGC

98251 CGAACAGGAA GCGGGCCGGC TGCACATCGA CCCGGTGCTG GTCGAGCCCG

98301 CGGGCGAGAC GATGACCGAG AGCGACACCG ACCGCATCGC TGCGGCGTTC

98351 GGCGCCAAGG TGCGCACGAT GTACAGCGCG ACCGAGTGCA CCTACCTCAG

98401 CCACGGCTGC GCCGAGGGCT GGTACCACGT CAACGACGAC TGGGCCGTGC

98451 TCGAACCGGT CGACGCCGAC CACCGGCCCA CCCCGCCGGG GGAGTTCTCG

98501 CACACCACCC TGATCAGCAA CCTCGCCAAC CGCGTCCAGC CGTTCCTCCG

98551 CTACGACCTG GGCGACAGCG TCATGCTCCG CCCCGACCCC TGCCCCTGCG

98601 GCACCCCCTC GCCCGCGATC CGGGTCCAGG GCAGGTCGGG CGACATCCTC

98651 ACCTTCCCCT CGGGCCGGGG CGACGACGTC AGCCTCGCCC CGCTCGCCTT

98701 CAGCAGCCTC TTCGACCGCA TGCCCGGAGT CGAGCTCTTC CAGATCGAGC

98751 AGACCGCGCC GTCGACCCTG CGCGTCCGCG TGGTCCAGGC GCCCGGCGCC

98801 GACGCGGACC ACGTGTGGCA GCGGGCCCAC GACGGGCTGA CCCACCTCCT

98851 CGCCGACAAC AAGCTCGACA ACGTAACCGT CGAACGGGGC GAGGAGCCGC

98901 CGCGGCAGGC ATCCGGCGGC AAGTACCGGA CGATCATCCC GCTCGCCGCC

98951 TGAACGCTCG CCGACTAGCC GCGCGCCGCC TGAGCTGCTC TCACCGCGCG

99001 TACGGGCGCA GCGGAGGCTC CTCGTCGACC CACGGCTGGC TGTGGATCAG

99051 CAGCTCGATC GGGAAGTTCA GCAGGCCGGG CAGGGCGTCG ACGGCCTCCT

99101 GGCTGTTGAG CGGCATGACC GGCTTGGCGC AGTGCGCGCG GTCGATGCGG

99151 CTCGTGTCGG CGGGACCGGG GTGCTCGATC GCATCGGCGA CCAGGTCGTA

99201 GCTGACCTGG TCGACGACCA TGGCGATGTG GGTCGGCCAC GGCCGACCCG

99251 GACAGATGTC CTGGACGCCG ATGCGGTGGG CGCCCGGCAG CGACGGCGCC

99301 TCCCCGTCGG CCACCACGGA CTCGTCGGCG TATGAATAGA TCGTGGTGTA

99351 CGACGGTCCT GCGGGCATGG GCGTGCCGTC GGCGCCCAGA GCCTTCGACC

99401 AGTTCGAGTC GCGGGCGAAC TGCAGGACCG ACGCCGGGCA GCCCGCCACC

99451 TCGGCGATCG GGCGGCAGGG CGAGGCCAGC CGGGTCCCCT GGAACGGGGA

99501 GCCCAGGGTC ACCATGTCGT CGACCTTCCC CGGCAGGTCC GGCCAGAAGC

99551 GCAGGGCCCA CGCCGTGAGG AGGCCGCCCT GGCTGTGCCC GACGAGATCG

99601 ACCTTCCGGC CGGTGGCCTC CTGGATCGCG CGGGTCGCGT ACACCACGTA

99651 CTCGACGGAC TCCTGCATGT CACGGAGCCC GCGACCGGGA GAATCCACCC

99701 AACAGGACTG GTAGCCCTTC TTCTTCAACT CGGCCATGTA GTTCCAGGCG

99751 TAGTTCTCCT CGCCCTTGAG GCCGGTCCCG GGCACGAAGA GGACGGTCGG

99801 CTTGTCACCG GCGTCACGCA GGTCCCCCAG CTCCGTCCCG CAGTGCAGCG

99851 CCTTGGCGAG CTCGGCCGCC GGTATCTCCA ACGGGGGAGA GGAAACATCC

99901 GCCGCCGAAG CGGCGGAGGC CGGAAGCACG GTGGCGGCCA GCACGGCCGC

99951 CACGAGTCCG CCGAGCCATG AGGACAAGCG CACGGTGACC TCCACAGGAA

100001 CCTTCACGAG TGAGCGGAAA CTCCCTCCGG AGGGAGCACC TCATCGTGCG

```
100051  GCGGCGCCAC AGTAGCCGTC AACTGCCCCA CGGGGCTGAG TAGTTGACAG
100101  TTGGCCGGGC TCGGCCGGCG AAGCGCCCGG GCCCCGCCGC CCCGCGCCGT
100151  GCGCGAGGGG TCCGTGACCT GGGTGGACGG TCCGGTTGGA CATCCCGGGG
100201  GAGCCTCTGG CATGGTCGCC CGTCCGTCCC CCTCAAGAAC CGAAGGGAGC
100251  GTCACGATCA CGATGATCGA AGTCAGCACG CGCAGCATGA AGGAAGCGGC
100301  TGCCGCCGAG CAGCTCCGCG CGGAGACCAC GACACTGGAC ATTCCAAAGG
100351  GTTTCGACCT GTGGACGGCC GACGAGATCG CGGAGTGGCT CGACGGCGTC
100401  GAGGACGACC CGGCAGTCTC CGACGCCGAC TTCTACGCGG CCCAGCAGCG
100451  GTGCGACGGG TCCTCGGCAC CGAGGGCACC TGACCCGCCG GCGGCCCTGC
100501  GCGGCCCTAC GTGTGCAGCG CCCCGTCCTC CTCCACATGC CCCTCCGGCT
100551  CCAGCTGGAT CGTCGAGTGG GCCACGTCGA AGTGGCCCCC GACACACCGC
100601  TGAAGGCGCC CCAGGAGCTC CCCGTACCCG CTCGCGAGAG CCTCCTCCGT
100651  GACCACCACG TGCCGCGGTGA GCACCGGCAT CCCCGAGGTG ACCGTCCAGC
100701  CGTGCAGATC GTGCACGGCG ACCACGCCCC GCTCCTCCAG CAGGTGCCGG
100751  CGCACCTCGC CGAGGTCGAC GTCCTGCGGG GTCGCCTCCA GCAGGACGTG
100801  CAAGGAGTCC CGCAGCAGGC CGTACGCGCG CGGCACGATC AGCAGGCCGA
100851  TGACGATCGA CGCGATCGGG TCGGCGGCCT GCCACCCCGT GAGCAGGATG
100901  ACCAGGCCGC CCACGATCAC CGCGACCGAG CCGAGCGCGT CGCCCAGCAC
100951  CTCCAGGTAC GCGCCCCGCA GATTGAGGCT CTTCTCCTTG GCGTCCCGCA
101001  GCAGCCACAG GCCCACCAGG TTGGCGGCGA GCCCGCCCAG CGCGACCACG
101051  AACATCAGGC CGCCCTTCAC CTCCACCGGC TCGCTGAACC GGCCGATCGC
101101  CGACCACAGG ACCCAGGCGA AGATGACGAC CAGGAGCAGC GCGTTCAGGA
101151  CCGCGGAGAA GATCTCCACG CGGTAGAACC CAAAGGTGCG CCGCGGCGTC
101201  GGCGCCCGCT GGGCGAGGGT GATGGCACCG AGGGCCAGCG AGACGCCGAC
101251  CGCGTCGGTC AGGCTGTGCG CGGCGTCGGC GAGCAGCGCG AGGCTGCCGG
101301  ACAGGAGCGC GCCGACCACC TGGATGACGG TGATCGAGCC GCTGATGCCG
```

```
101351  ATGGTCCACA GCAGGCGCTT GCGGTACGTG CCGCTGAGAG TGCCGCCCGC
101401  CGCCCCGGCG GACGGACCGT GGTCGTGCCC CATGCCCGCG AGTGGACCAC
101451  GGCGGCGCGG CACCCGCCAC CGAGCGGCCG CCGGTCGGCT CAGTGCAGCC
101501  GGGCCTGGGT GGAGGTGTCG CGCTGGTGCG GGATGCCGAG CGGCGGCGGC
101551  AGCTCGCCCT GCTGCACCCT GACCGTGCGC ACGGGGGCGG GGACCCGGAT
101601  GCCCTCGGCG CGGTAACGCT GGTGCAGGCG CTTGATGAAC TCGTGCTTGA
101651  TGCGGTACTG GTCGCTGAAC TCGCCGACGC CGAGGATCAC CGTGAAGCTG
101701  ATCCGCGAGT CGCCGAAGGT GTGGAAGCGG ATCGCCGCCT CGTGGTCGGG
101751  GACCGCGCCG GTGATCTCGG CCATCACCTC GTCCACCACC TCGGTCGTGA
101801  CCTTCTCGAC CTGCTCCAGG TCGCTGTCGT AGCTGACCCC GACCTGCACC
101851  ATGATCGACA GCTCCTGCTC GGGGCGGCTG TAGTTGGTCA TGTTGGTGCC
101901  GGCGAGCTTC GCGTTGGGGA TGATGACGAG GTTGTTGGAG AGCTGGCGGA
101951  CCGTGGTGTT GCGCCAGTTG ATGTCGACGA CGTAGCCCTC CTCCCCGCTG
102001  CTGAGCTGGA TGTAGTCGCC GGGCTGCACG GTCTTCGCGG CGAGGATGTG
102051  CACGCCCGCG AAGAGATTGG CGAGCGTGTC CTGCAGTGCG AGGGCGACCG
102101  CGAGACCTCC CACGCCGAGG GCGGTGAGCA GCGGTGCGAT GGAGATGCCG
102151  AGGGTCTGAA GGACGATGAG GAAGCCCATC GCGAGCACCA CGACGCGGGT
102201  GATGTTCACG AAGATGGTGG CCGATCCGGC CACTCCGGAG CGGGACTGTG
102251  CCACGGCCTT CACCAGGCCG GTGACGATCC GGGCCGCCGT GAGCGTGGCG
102301  GCCAGGATGA GCAGCGCGGT CAGCGTCATG GTGACGTTGC GTCCGGTGCG
102351  CGGCGTGAGC GGCAGCGCGC CCGCCGCGGC GGCGAGCCCG GCGGTGATGG
102401  CCGCGCAGGG CACGAGGGTG CGCAGGGCGT CGACGATGAC GTCGTCACCG
102451  CTCCACCGGG TTTTGCTCGC CCGTTCGCCG AGCCACCTCA GAAGTGCGCG
102501  GAGCAGCAGC CCGGCGACGA CGCCGGCGAC GACCGCGATA CCGGCCACGA
102551  TCCAGTCGTG CAGTGTGAGG GCACGGGTCA TCAGTTCGCT CCCGTCGTAC
102601  GGGGGGAGTG CGCCTGTGTG GGGCGTATGT GATGTGACGT CACCTTGTGA
```

```
102651  TACCTGCTCG ATTCCGGGGA GTGCGGTCAC GCCGGGACGA GAGCTCGGTT
102701  CCGGCGCGGA CGTCATCCTG CCCCATCCGC CCACGGCAGG CGTGCATACC
102751  CCCACCTGGA TCTTCACAGA CCGGCCACGT CTGTCCATGC GCCGATGAGC
102801  GCGCTGCCCG TGGTAAAGCA TTGAGTCAGG CGATTTGGCC ACTCGGCACT
102851  CGGCGGACCG GTCGAGCCGG TCGATCTACG TGAGCGGAGG CGGTTGAGCA
102901  TGGCGTCCAT GTGCAGACCC GGAATGTCAC CCGTCAATTC GCACAACGAG
102951  TGGGATCCGC TGGAGGAGAT CATCGTCGGG CGGCTGGAGG GCGCGACCAT
103001  TCCCTCCAGC CATCCGGTCG TGGCGTGCAA CATCCCGACC TGGGCGGCAC
103051  GGCTGCAGGG TCTCGCCGCC GGGTTCGAGT ATCCGCAGCG GCTGATCGAG
103101  CCGGCGCAGC AGGAGCTCGA CCAGTTCATC GCTCTCCTGC AATCCCTCGA
103151  CGTCACAGTG AGACGGCCGG CGGCCGTCGA CCACAAGCAC CGCTTCGGGA
103201  CCCCCGACTG GCAGTCGCGC GGCTTCTGCA ATTCCTGTCC GCGGGACAGC
103251  ATGCTCGTCG TCGGCGACGA GATCATCGAG ACCCCGATGG CGTGGCCGTG
103301  CCGCTGTTTC GAGACGCACT CGTACCGCGA ACTCCTCAAG GACTACTTCC
103351  GGCGCGGCGC GCGCTGGACG GCGGCGCCGC GCCCCCAGCT CACCGAGGCC
103401  CTGTACGAGA AGGACTTCCG CCCTCCCGAG GAGGGCGAAC GATGCGCTAC
103451  ATCCTCACCG AGTTCGAGCC GGTGTTCGAC GCGGCGGATT TCGTGCGGGC
103501  GGGCCGCGAC CTGTTCGTGA CGCGGAGCAA CGTCGCCAAC CTGCTGGGCA
103551  TCGAGTGGCT GCGCCGCCAC CTTCGGGCCG GAGTACCGCG TGCCACGAGA
```

DECLARATION, POWER OF ATTORNEY AND POWER TO INSPECT

As a below named inventor, I hereby declare:

that my residence, post office address and citizenship are as stated below next to my name;

that I verily believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the invention entitled: **POLYKETIDES AND THEIR SYNTHESIS** the specification of which [check one(s) applicable]

- X was filed 30 May 2000 as International Patent Application No. PCT/GB00/02072, on which U.S. National Stage Application No. 09/980,217 is based; and/or
- ___ was amended by Amendment filed _____ (if applicable), and/or
- ___ is attached to this Declaration, Power of Attorney and Power to Inspect;

that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above; and

that I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Rule 56(a) [37 C.F.R. §1.56(a)]

CLAIM UNDER 35 U.S.C. §119: I hereby claim foreign priority benefits under 35 U.S.C. §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application of which priority is claimed:

Prior Foreign Application(s) Appln No.	Country	Filing Date Day-Mon-Year	Priority Claimed Yes - No
9912563.5	Great Britain	28-05-1999	Yes

POWER OF ATTORNEY: As inventor, I hereby appoint **DANN, DORFMAN, HERRELL AND SKILLMAN, P.C.** of Philadelphia, Pennsylvania, and the following individual(s) as my attorneys or agents with full power of substitution to prosecute this application and to transact all business in the United States Patent and Trademark Office connected therewith: **Patrick J. Hagan, Reg. No. 27,643 and Kathleen D. Rigaut, Ph.D., Reg. 43,047.**

POWER TO INSPECT: I hereby give **DANN, DORFMAN, HERRELL AND SKILLMAN, P.C.** of Philadelphia, Pennsylvania or its duly accredited representatives power to inspect and obtain copies of the papers on file relating to this application.

SEND CORRESPONDENCE TO: **CUSTOMER NUMBER 000110**

DIRECT INQUIRIES TO: **Telephone: (215) 563-4100**
Facsimile: (215) 563-4044

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

SOLE OR FIRST JOINT INVENTOR

Full Name Peter (First) Francis (Middle) Leadlay (Last)

Signature PFLeadlay

Date 4 April 2002

Residence Cambridge, U.K.
City State or Country

Citizenship GB GBX

Post Office Address:
6 Westberry Court, Pinehurst,
Street Address
Grange Rd, Cambridge UK CB3 9BG
City State or Country Zip Code

SECOND JOINT INVENTOR (IF ANY)

Full Name James (First) Staunton (Middle) (Last)

Signature JStaunton

Date 8 April 2002

Residence Cambridge UK
City State or Country

Citizenship GB GBX

Post Office Address:
29 Porson Road
Street Address
Cambridge CB2 2ET
City State or Country Zip Code

[redacted]

THIRD JOINT INVENTOR (IF ANY)

Full Name Mark YAN Oliynyk
First Middle Last

Signature [signature]

Date 4th April 2002

Residence CAMBRIDGE, UK GB
City State or Country

Citizenship UKRAINIAN

Post Office Address:

61 ST THOMAS'S SQUARE
Street Address

CAMBRIDGE, UK CB1 3TG
City State or Country Zip Code

FOURTH JOINT INVENTOR (IF ANY)

Full Name _____
First Middle Last

Signature _____

Date _____

Residence _____
City State or Country

Citizenship _____

Post Office Address:

_____
Street Address

_____
City State or Country Zip Code

[redacted]

[redacted]